Development of new classification models based on Raman spectroscopy and MALDI spectrometry as novel tools for liver cancer diagnostic

Dissertation

zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)

seit 1558

vorgelegt dem Rat der Medizinischen Fakultät
der Friedrich-Schiller-Universität Jena

von Dipl.-Biol. Tatiana Tolstik

geboren am 06.07.1987 in Minsk, Weißrussland

Gutachter

1. Prof. Dr. Andreas Stallmach

2. Prof. Dr. Jürgen Popp

3. Prof. Dr. Andrea Tannapfel

Tag der öffentlichen Verteidigung: 07.06.2016

For my family

“Be realistic: Plan for a miracle”
Osho

Content

List of abbreviations

AASLD	American Association for the Study of Liver Diseases
CARS	Coherent anti-Stokes Raman spectroscopy
CCA	Cholangiocellular carcinoma
CCD	Charge-coupled device
CLT	Cadaveric liver transplantation
CT	X-ray computed tomography
EASL	European Association for the Study of the Liver
EORTC	European Organisation for Research and Treatment of Cancer
GPC3	Glypican 3
GS	Glutamine synthetase
HBV	Hepatitis B virus
HCA	Hierarchical cluster analysis
HCC	Heptocellular carcinoma
HCV	Hepatitis C virus
H&E	Hematoxylin and eosin stain
HSP-70	Heat shock protein 70
IARC	International Agency for Research on Cancer
LDLT	Live donor liver transplantation
LIF	Laser-induced fluorescence
LTRS	Laser tweezers Raman spectroscopy
m/z	Mass-to-charge ratio
MALDI IMS	Matrix-assisted laser desorption/ionization imaging mass spectrometry

MRI	Magnetic resonance imaging
N-FINDER	N-finder algorithm
PEI	Percutaneous ethanol injection
RF	Random forest
RFA	Radiofrequency ablation
SERS	Surface-enhanced Raman spectroscopy
SVM	Support vector machine
TACE	Transarterial chemoembolization
TAE	Transarterial embolization
TOF	Time of flight
Ubi	Ubiquitin
WHO	World Health Organization

Summary

Hepatocellular carcinoma (HCC), as the most common type of primary liver tumor, is the third leading causes of cancer mortality worldwide. Usually occurring along with liver fibrosis or cirrhosis together with atypical vascularisation pattern, HCC is often hard to diagnose. The pathological examination of cytohistological samples is a crucial step in the diagnosis of liver cancer that can be a difficult task even for an experienced pathologist. Currently classical and immunohistochemical staining techniques as well as gene expression profiles are common in clinical practice for the investigation of biopsies with suspected HCC. Nevertheless, morphological features have not been proven to be clinically significant. Therefore, novel optical diagnostic tools are currently explored that are able to detect various biomarkers in cytological and histological samples. In recent years, a wide range of optical and spectroscopic techniques, e.g. Raman Spectroscopy and Matrix-Assisted Laser Desorption/Ionization imaging mass spectrometry (MALDI IMS) have been applied and proven to be reliable methods for cancer diagnostics.

The objective of the present thesis was to investigate the applicability of Raman spectroscopic imaging and MALDI IMS as diagnostic tools for liver cancer detection on the cellular and tissue levels. In the first study cells from different liver cancer cell lines (HepG2 and SK-Hep1) were analyzed by mean of Raman imaging in order to develop new algorithms for prediction of unknown cells from different origin. A support vector machine based classification model resulted in a prediction accuracy of 93 % for prediction of HepG2 and SK-Hep1 cells. By applying hierarchical cluster analysis to Raman maps of each single cell, different cellular compartments such as nucleus, lipid droplets and cytoplasm were differentiated and analyzed separately. Using only spectral information of cytoplasmic lipids, the prediction accuracy of support vector machine classification model improved to 96 % in comparison to 91 % for nuclei, 87 % for cytoplasm and 93 % for the complete cell information. Furthermore, the metabolic changes during different proliferation stages of HCC cells based on Raman spectra were investigated. The data showed high expression of unsaturated fatty acids during the proliferation in comparison with apoptosis. Differences in the Raman spectra allowed us to classify and predict the proliferation behavior of the cells with an accuracy of 99 %.

To investigate the diagnostic capacities of Raman spectroscopy on tissue level, predefined malignant and non-malignant regions from patient samples with HCC were analyzed. We were able to detect lipids, proteins, collagen and cholesterol ester as separate components within the Raman maps and analyzed their distribution in different hepatic lesions. A random forest classification model allowed us to predict HCC tissue regions with sensitivity of 76 % and specificity of 93 %. The main Raman bands important for the model were identified as fatty acid bands and especially attributed to palmitic acid, known to play an important role in the HCC development. Furthermore, the developed classifier was used in order to predict tumor margins between HCC tissue and surrounding fibrosis. Consequently, Raman spectroscopy showed a high potential for surgical margin status detection, especially tumor margins.

As a second diagnostic technique for analyses of localization of different proteins within malignant and non-malignant tissue regions MALDI IMS was applied. Four proteins (6274, 6647, 6222 and 6853 m/z) with significantly higher expression profile in the HCC tissue regions in comparison with non-tumorous liver tissue were identified. The developed classification model, based on three most significant proteins, allowed prediction of HCC with sensitivity and specificity of 90 %.

In conclusion, the obtained results showed that Raman and MALDI IMS imaging techniques can successfully detect and localize a variety of molecules within a sample. We proved that Raman spectroscopy can effectively classify and predict molecular changes in liver carcinogenesis on cellular and tissue levels in label-free manner. Furthermore, an application of MALDI IMS provides a specific proteomic signature that can be used as biomarkers for diagnostic of liver cancer.

Zusammenfassung

Die häufigste Ursache eines primären Lebertumors ist das Hepatozelluläre Karzinom (HCC) und dies ist damit weltweit die dritthäufigste Krebstodesursache. In der Regel geht das Karzinom gleichzeitig mit einer Leber-Fibrose oder –Zirrhose und mit untypischen Vaskularisationen einher, wodurch sich ein HCC häufig schwer diagnostizieren lässt. Die pathologische Untersuchung der zytohistologischen Probe ist der entscheidende Schritt für die Diagnose von Leberkrebs. Dies kann selbst für einen erfahrenen Pathologen eine schwierige Aufgabe darstellen. In der klinischen Praxis werden gegenwärtig klassische und immunhistochemische Färbemethoden sowie Genexpressionsprofile zur Untersuchung von Biopsien mit Verdacht auf HCC eingesetzt. Dennoch wurden morphologische Besonderheiten nicht auf ihre klinische Signifikanz geprüft. Daher werden gegenwärtig neuartige optische Diagnosegeräte erforscht, welche in der Lage sein sollen, verschiedene Biomarker in zytologischen und histologischen Proben zu detektieren. In den letzten Jahren wurde eine Vielzahl an optischen und spektroskopischen Techniken zur Krebsdiagnostik eingesetzt und auf deren Methodenzuverlässigkeit getestet, z. B. Raman-Spektroskopie und Matrix-unterstützte Laser-Desorption/Ionisation Massenspektrometrie (MALDI IMS).

Ziel der vorliegenden Dissertation, war die Erforschung von Raman-Spektroskopie und MALDI IMS als geeignetes Diagnose-Instrument zur Detektion von Leberkrebs auf zellulärer- und Gewebe-Ebene. In der ersten Untersuchung, wurden Zellen unterschiedlicher Leberkrebs-Zelllinien (HepG2 und SK-Hep1) mittels bildgebender Raman-Spektroskopie analysiert, um neue Algorithmen zur Voraussage unbekannter Zellen aus abweichenden Quellen zu entwickeln. Ein Support Vector Machine basierendes Klassifikationsmodell ergab eine 93 %ige Vorhersagegenauigkeit für die Unterscheidung von HepG2 und SK-Hep1. Unter Verwendung einer hierarchischen Clusteranalyse auf die einzelnen Raman-Abbildungen jeder Zelle, wurden verschiedene zelluläre Kompartimente, wie Nukleus, Lipidtropfen und Zytoplasma, differenziert und separat ausgewertet. Wurden ausschließlich spektrale Informationen von zytoplasmatischen Lipiden verwendet, verbesserte sich die Vorhersagegenauigkeit des Klassifikationsmodells auf 96 % im Vergleich zu Nuklei mit 91 %, Zytoplasma mit 87 % und der gesamten Zellinformation mit 93 %. Zusätzlich wurden die metabolischen Veränderungen der HCC-Zellen anhand von Raman-Spektren während verschiedener Proliferationsphasen erforscht. Gezeigt wurden erhöhte Expressionswerte von

ungesättigten Fettsäuren während der Proliferation, im Vergleich zur Apoptose. Abweichungen in den Raman-Spektren erlaubten uns das Proliferationsverhalten der Zellen mit einer Genauigkeit von 99 % zu klassifizieren und vorherzusagen.

Um die diagnostischen Kapazitäten der Raman-Spektroskopie auf der Gewebe-Ebene zu untersuchen, wurden vordefinierte maligne und nicht maligne Bereiche der Patientenproben mit dem HCC analysiert. Wir waren imstande, Lipide, Proteine, Kollagen und Cholesterinester als separate Komponenten innerhalb der Raman-Abbildungen zu detektieren und deren Verteilung in verschiedenen hepatischen Läsionen auszuwerten. Ein Random Forest Klassifikationsmodel erlaubte uns, die HCC-Geweberegionen mit einer Sensitivität von 76 % und einer Spezifizität von 93 % vorherzusagen. Die wichtigsten und für das Model bedeutendsten Raman-Banden wurden als Fettsäure-Banden, insbesondere Palmitinsäure, identifiziert. Von diesen ist bekannt, dass sie eine bedeutende Rolle bei der Entstehung von HCC haben. Des Weiteren wurde der entwickelte Klassifizierer genutzt, um Tumorgrenzen zwischen HCC-Gewebe und der umgebenden Fibrose vorauszusagen. Demzufolge weist die Raman-Spektroskopie ein großes Potential zur operativen Feststellung von Grenzzuständen auf, insbesondere von Tumorgrenzen.

Als zweite diagnostische Analysetechnik wurde die MALDI IMS zur Lokalisation verschiedener Proteine innerhalb maligner und nicht maligner Geweberegionen angewandt. In der HCC-Geweberegion wurden im Vergleich zur Region ohne Lebertumor vier Proteine (6274, 6647, 6222 und 6853 m/z) mit signifikant erhöhtem Expressionsprofil identifiziert. Das auf die drei bedeutendsten Proteine entwickelte Klassifikationsmodell, erlaubte die Vorhersage von HCC mit einer Sensitivität und Spezifizität von 90 %.

Zusammenfassend konnte gezeigt werden, dass die erhaltenen Ergebnisse der bildgebenden Raman und MALDI IMS Technik erfolgreich zur Detektion und Lokalisation einer Vielzahl von Molekülen innerhalb einer Probe führten. Wir belegten, dass die Raman-Spektroskopie markierungsfrei molekulare Veränderungen in der Leber-Karzinogenese auf zellulärer und Gewebe-Ebene effektiv einordnen und voraussagen konnte. Zusätzlich lieferte die Anwendung von MALDI IMS spezifische Proteom-Signaturen, welche als Biomarker zur Leberkrebs-Diagnostik genutzt werden können.

1. Introduction

1.1 Liver cancer and hepatocellular carcinoma

1.1.1 Definition of liver cancer

The earliest description of the primary liver cancer as an independent disease was given by C.A. Rokitansky in 1849 (Kuntz and Kuntz, 2008). Thereafter, the existence of both primary and secondary (metastatic) liver tumors was accepted. Today primary liver carcinoma is the sixth most common cancer type with more than 749 000 new cases per year and is the third leading causes of cancer mortality worldwide (EORTC, 2012; IARC and WHO, 2014). Hepatocellular carcinoma (HCC) and Cholangiocellular carcinoma (CCA) are the most common types of primary liver cancer with an occurrence rate of 90 % and 10 %, respectively (EASL and EORTC, 2012; Kuntz and Kuntz, 2008).

1.1.2 Hepatocellular carcinoma

Epidemiology and risk factors

HCC is a cancer that originates in liver cells (hepatocytes) and accounts for more than 692 000 cases per year (Behne and Copur, 2012). Because of the poor prognosis, the number of HCC related deaths is almost the same: 589 000 (Parkin et al., 2005). This type of liver cancer affects persons of all age groups, but is more common in elderly (older than 70 years), and there is a higher incidence among men than women (2.4 : 1) (EASL and EORTC, 2012; Kuntz and Kuntz, 2008). The majority of cases are associated with liver cirrhosis, which makes early diagnosis of HCC complicated.

The occurrence rate of this disease varies with geographical location and correlates with distribution of Hepatitis B virus (HBV) or Hepatitis C virus (HCV) infections (Kuntz and Kuntz, 2008). The high incidence of HCC (85 % of all cases) observed in sub-Saharan Africa, East Asia and Melanesia, whereas the incidence is low in developed areas except for Southern Europe (10.5 per 100 000 inhabitants) (EASL and EORTC, 2012). The main risk factors for development of HCC in Asia and Africa are HBV (70 %), HCV (20 %) and aflatoxins (10 %); and in Europe and the USA: HCV (ca. 60 %), alcohol (20 %) and HVB (ca. 15 %) (EASL and EORTC, 2012; Kuntz and Kuntz, 2008).

Additional factors influencing development of HCC can be seen in Figure 1 (Bergk and Schott, 2011).

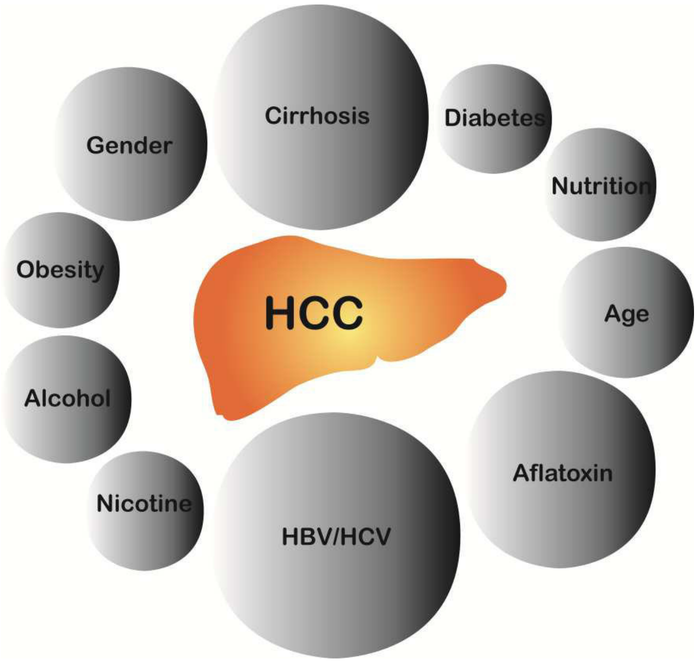


Figure 1. Risk factors for HCC development. The size of the circles correlated with the significance for each risk factor. Adapted from Bergk and Schott (Bergk and Schott, 2011).

Symptoms and diagnostic

Often, patients with suspected HCC admitted with symptoms of underlying liver diseases (Sugano et al., 1994). In about 70-80 % of HCC cases the disease is coexisting with liver cirrhosis, that often results in complications such as encephalopathy, ascites, jaundice or bleeding (EASL and EORTC, 2012; Sugano et al., 1994; Zhu et al., 2013). The main symptoms of HCC are abdominal pain, bloating, flatulence, loss of appetite, nausea, weight loss, fatigue, weakness and stool irregulation (Kuntz and Kuntz, 2008).

Until 2000 the diagnostics of HCC was based on biopsy (EASL and EORTC, 2012). Due to the risk of complications and the complexity to differentiate between early HCCs and high grade dysplastic nodules, the non-invasive criteria was developed in 2001 by European Association for the Study of the Liver (EASL). This new criteria is based on a combination of imaging (4-phase computed tomography (CT) and dynamic contrast enhanced magnetic resonance imaging (MRI)) and laboratory findings (Bruix et al., 2001; EASL and EORTC, 2012). In 2005, additions concerning a new HCC radiological hallmark for diagnosis of HCC were included by EASL and the American Association for the Study of Liver Diseases (AASLD) (Bruix et al., 2001). At present the guidelines from AASLD and EASL require for both types of nodules (of 1-2 cm and above 2 cm) presence of HCC radiological hallmarks in the imaging techniques CT or MRI (EASL and EORTC, 2012). Nevertheless, in case of unclear findings the implementation of second medical imaging technique is necessary.

Such an imaging technique as contrast enhanced ultrasonography is not recommended for diagnostic purposes due to lack of specificity for HCC (Halperin et al., 2013). Because of often inadequate sampling, liver biopsy, as another diagnostic method, also provide a high false negative rate (up to 30%) in diagnostics of HCC (Forner et al., 2008). Moreover, pathological diagnostics of liver biopsy can often be challenging even for an expert in pathology, especially in case when the nodules ranges from 1 to 2 cm (Roskams and Kojiro, 2010). Nevertheless, liver biopsy sampling is recommended when the diagnostics of HCC cannot be made based on non-invasive criteria because of abnormalities in radiological examinations (EASL and EORTC, 2012). Thus, biomarkers can improve the examination of the tissue specimens by immunohistochemistry. The most promising tissue markers for HCC are glypican 3 (GPC3), heat shock protein 70 (HSP-70), glutamine synthetase (GS) and/or expression of *GPC3*, *lymphatic vessel endothelial hyaluronan receptor* and *surviving* genes (EASL and EORTC, 2012). However, even complex staining of GPC3 in combination with HSP-70 and GS can provide sensitivity only of 72 % and specificity of 100 % for HCC diagnosis.

Therapy and prognosis

Various treatment options can be offered to the patients diagnosed with HCC, including liver resection, transplantation, ethanol injection, radiofrequency ablation, transarterial embolization and etc. (Table 1) (EASL and EORTC, 2012). The choice of therapy depends on the stage, size and number of nodules, as well as on the increased portal pressure and elevated bilirubin level, existence of tumor capsule, extrahepatic metastases or associated diseases.

Table 1. Algorithm for staging and treating patients diagnosed with HCC. Adapted from Roskams and Kojiro (Roskams and Kojiro, 2010).

Resection	➢ Very early stage, single nodule < 2 cm, carcinoma *in situ* ➢ Early stage, single nodule <5 cm, no increased portal pressure or elevated bilirubin levels	**5-year survival** **40-70 %**
Liver transplantation (CLT or LDLT)	➢ Early stage, single nodule <5 cm, increased portal pressure and elevated bilirubin levels, no associated diseases ➢ Early stage, 3 nodules ≤3 cm, no associated diseases	
PEI or RFA	➢ Early stage, 3 nodules ≤3 cm, associated diseases	
TAE or TACE	➢ Intermediate stage	**Median survival** **11-20 months**
Kinase inhibitor drugs (i.e Sorafenib and Doxorubicin)	➢ Advanced stage	
Supportive care	➢ Terminal stage	**Survival** **< 3 months**

CLT - cadaveric liver transplantation, LDLT - live donor liver transplantation, PEI - percutaneous ethanol injection, RFA - radiofrequency ablation, TAE - transarterial embolization, TACE - transarterial chemoembolization.

Early diagnostics of HCC, when the tumor is around 2 cm in size, allow offering all treatment options to the patient. However HCC is often diagnosed at the late stage when the disease is in an untreatable condition (Manghisi et al., 1998). Then estimated survival of the patient is from 6 to 20 months. Thus, the surveillance strategies and early diagnosis of the liver tumor are crucial for potential treatment.

The median survival in patients with HCC received transarterial chemoembolisation (TACE) is around 16 months, for the patients treated with percutaneous ethanol injections (PEI) is around 11 months and for the patients receiving TACE followed by PEI is around 24 month (Greten et al., 2005). Therefore, the only therapies promising good prospects of a cure are liver resection and transplantation, which provides a 5-year survival rate of 60-80 % (Llovet et al., 2005; Mazzaferro et al., 1996). While the liver is one of two organs in the human body, next to the skin, capable of regeneration, the first-line treatment for the patients with small single tumor and sufficient liver function is tumor resection. This type of treatment provides a surgical removal of the tumor and surrounding tissue and preserves a part of the liver sufficient for body function. That procedure can be high mortal for cirrhotic patients. In that case more than 40 % of the liver needs to be remaining and for non-cirrhotic patients 25 % of liver remnant is enough. Surgical resection offers the best prognosis for long-term survival, but is recommended only to 40 % of the patients in Asia and to 5 % in the West (Belghiti et al., 2000; Lang et al., 2005). In cases where the patient is not suitable for resection, liver transplantation should be offered as next treatment option in accordance with the Milan criteria, that were developed for selecting patients for liver transplantation (Mazzaferro et al., 1996). Liver transplantation is the surgical replacement of a diseased liver with a healthy donor liver (orthotopic transplantation) or a part of the liver (living donor liver transplantation) (EASL and EORTC, 2012). According to EASL clinical practice guidelines there is a 5-year survival rate of 65-78 % of the patients and 10-year survival rates of around 50 % for the patients after liver transplantation.

1.2 Raman Imaging Spectroscopy

1.2.1 Theoretical basis of Raman spectroscopy

During a trip from India to England in 1921 the Indian physicist C.V. Raman performed on-board experiments that showed that the color of the sea was a result of direct scattering of light and independent from reflection or absorption of light from the sky (Ellis et al., 2013). Later this work was submitted to *Nature* with the title "The color of the sea" and refuted the theory suggested by Lord Rayleight that the sea color was just a reflection of the color of the sky. After this investigation, C.V. Raman and collaborators focused on this new research field in physics; scattering of the light in liquids and solids. Thereafter C.V. Raman discovered an inelastic scattering effect in 1928 named after him and got the Nobel prize for Physics in 1930. Despite the Raman effect was described theoretically by Adolf Smekal already in 1923, it was also observed by C.V. Raman and K.S. Krishnan in liquids and by G. Landsberg and L.I. Mandelstam in crystals in 1928 (Landsberg and Mandelstam, 1928; Raman and Krishnan, 1928; Smekal, 1923).

Enter into details, the scattering or absorption of the light happens when a photon of light impact on a particle (a molecule or an atom) in the ground state (Dieing et al., 2011). If there is no energy transfer between the photon and the particle, e.g. due to the smaller molecule size compared to the radiation wavelength, the effect is called elastic or Rayleigh scattering. This scattering will not change the material, hence the frequencies of the incident and scattered radiation will remain the same. Nevertheless, by interacting with the molecule, individual atoms or groups of atoms of which can oscillate relative to the other part of the molecule, the emitted radiation will possess long-wave or short-wave frequency shifts with respect to the initial excitation radiation at a frequency ω. Thus, scattering with the shifted to lower frequencies (ω-Ω) corresponds to the case of a Stokes Raman scattering, while shifted to higher frequencies (ω+Ω) gives an anti-Stokes Raman scattering, where Ω is vibrational frequency of molecule. These three effects can be displayed via Jablonski diagram (see Figure 2).

Raman spectroscopy is an optical approach utilizing Raman scattering for the analysis of substances. This technique provides information on molecular level by investigating functional groups, bonding types and molecular conformation in a sample (Schmitt and

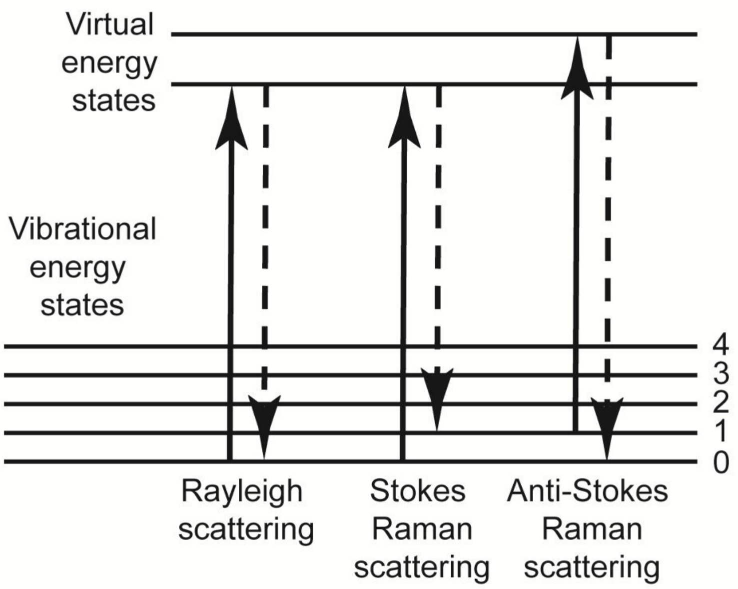


Figure 2. Schematic representation of the Jablonski diagram. Adapted from Pavlina (User:Pavlina2.0, 2009)).

Popp, 2006). The spectra of Raman scattered light are unique fingerprints with specific spectral bands providing direct information about functional groups, bonding types and molecular conformation of solid, liquid or gaseous materials. A meanwhile relatively straightforward technical setup and great molecular sensitivity has led to the fact that Raman spectroscopy has become a powerful tool not only for chemical analysis but also in other research fields. The implementation of Raman spectroscopy in the Mars mission of European Space Agency and National Aeronautics and Space Administration is a good example for its wide application. And what is of high interest, Raman spectroscopy has recently made significant progress in the field of biomedicine (Movasaghi et al., 2007).

1.2.2 Biomedical applications of Raman spectroscopy

As it was outlined above, Raman spectroscopy is a non-destructive and label-free technique for *in situ* investigation of molecular information in a sample. Beginning with applications in physics and chemistry Raman spectroscopy extended up to biomedical

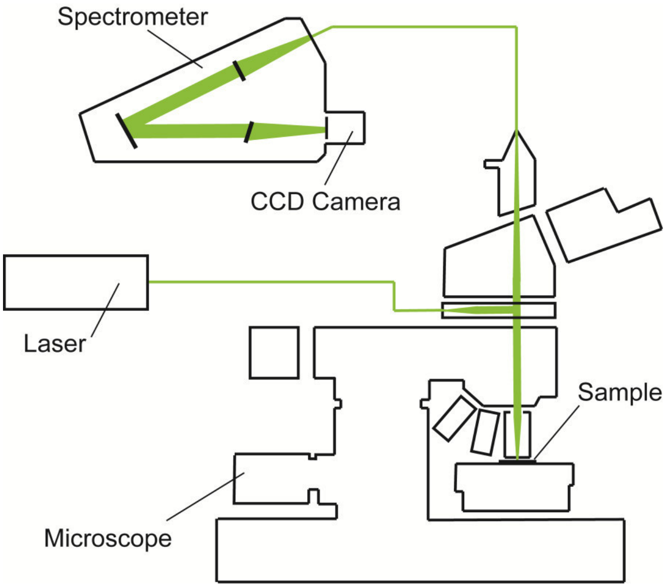


Figure 3. Schematic layout of a Raman microscope. Adapted from Dieing *et al.* (Dieing et al., 2011).

science revealing studies of cell identification, as well as molecular signatures of diseases and individual bacteria (Bielecki et al., 2012; Kloss et al., 2013; Neugebauer et al., 2010a). Furthermore, recent developments in optics allow a combination of Raman spectroscopy and high resolution confocal microscopy that provides a new non-invasive method to image a sample. Figure 3 shows a typical Raman microscope system including a spectrometer with a charge-coupled device (CCD) detector, an excitation laser and a microscope body. The possibility to image a sample point by point is provided by a piezo-driven stage. The optimization of the throughput of the Raman microscope is crucial due to the weak Raman signal (approximately 1 of 10 million of emitted photons) (Dieing et al., 2011).

Because of the submicron resolution (<1 µm) Raman spectroscopy in combination with confocal microscopy allows imaging of cellular organelles, including nucleus, lipid droplets, nucleoli, mitochondria and cytoplasm without use of fluorescent dyes. Diversity in the cell morphology and organelle composition allows identification of different cell types by Raman imaging spectroscopy (Bocklitz et al., 2009; Chan and Lieu, 2009; Chan et al., 2009; Ghita et al., 2012; Harris et al., 2009; Harz et al., 2008; Hawi et al., 1996; Hedegaard et al., 2010; Jess et al., 2007; Lerou and Daley, 2005; McNeish, 2004; Neugebauer et al., 2010b; Oshima et al., 2010; Schulze et al., 2013; Tan et al., 2012; Yu et al., 2006). Time dependent changes of the organelles in size, amount, shape or composition as a response to stress factors can be successfully used as differentiating criteria (Guarracino et al., 2011; Kourtis and Tavernarakis, 2009; Krafft et al., 2006; Kroemer et al., 2010; Short et al., 2005; Verrier et al., 2004). Raman bands associated with the nucleus allow investigating changes during cell cycle, proliferation stages and apoptosis of the cells (Guarracino et al., 2011; Huser et al., 2009; Krafft et al., 2006; Matthäus et al., 2006; Short et al., 2005; Verrier et al., 2004). In addition, treatment effectiveness and drug delivery which are of high interest for pharmaceutical science have been monitored successfully by Raman spectroscopy as a label-free and non-invasive technique in recent years (Wilson and Jordan, 1995; Zoladek et al., 2011).

Apart from cell studies, the use of Raman spectroscopy in tissue diagnostics has a big potential based on developments of fiber optic probes that will allow the use of this method during endoscopy in clinical routine in future. The submicron resolution of the Raman imaging system allows detection of different biochemical compounds in tissue sections. The correlation of the obtained Raman spectral information with tissue morphology after subsequent histological staining, which is currently a gold standard method for cancer diagnostics in histopathology, can provide additional diagnostic value. Nowadays, Raman imaging has already been successfully applied as an *ex vivo* diagnostic method to investigate colon, cervical, brain, oral and stomach malignant tissues (Beljebbar et al., 2009; Bergholt et al., 2010; Bergner et al., 2012; Guze et al., 2014; Kamemoto et al., 2010; Lloyd et al., 2012; Mavarani et al., 2013; Patel et al., 2011). A special attention is paid on the application of fiber optic probes for *in vivo* diagnostics. The use of Raman probes in bronchoscopy, endoscopy of the upper gastrointestinal tract and miniaturized Raman endoscope in mice have demonstrated the proof of principle for *in vivo* studies (Bergholt et al., 2010; Short et al., 2008; Taketani et al., 2013).

Based on data from The World Health Organization 216 infectious diseases caused by viruses, bacteria and macro-parasites leads to 25 % of all deaths worldwide (IARC and WHO, 2014). Due to the submicron resolution Raman spectroscopy is also suitable for identification of pathogens (Kloss et al., 2013; Schröder et al., 2013). Moreover, Raman spectroscopic investigations of malaria, meningitis, chlamydia infections and sepsis show its applicability in microbiology (Asghari-Khiavi et al., 2011; Harz et al., 2008; Kang et al., 2011; Neugebauer et al., 2014; Puskar et al., 2007; Szaszák et al., 2013).

During the last decade, Raman spectroscopic imaging found a variety of applications in biomedicine (as described in chapter 4.1) that proves the possibility of its future use as a diagnostic tool in the clinic. Nevertheless, there are both advantages and disadvantages of this method that can be seen in Table 2.

Table 2. Advantages and disadvantages of using Raman spectroscopy for biomedical applications

Advantages	Disadvantages
+ Non-destructive + Label-free + High resolution + Composite “fingerprint” spectral information + Simple sample preparation and small sample amount + Weak water scattering + Suitable for fiber optic probes	- Weak Raman signal - Autofluorescence - Lack of sensitivity for low concentrations - Chemometric data analysis is necessary - Long measurement time - Individual molecular component identification may not be directly observed

1.3 MALDI imaging mass spectrometry

Matrix-assisted laser desorption/ionization imaging mass spectrometry (MALDI IMS) is the modern technique that combine highly selective and sensitive mass spectrometry with spatial analysis, allowing visualization of a large amount of analytes (proteins, peptides, lipids and other small molecules) within a sample (Dudley, 2014). By been originally focused on the analysis of intact proteins and peptides, MALDI IMS has found nowadays various applications in the analysis of lipids, genotypes and micro-organisms as well as discovery of biomarker and metabolomics (Wiley and McLaren, 1955).

After the ionization process was described in 1980s, the development of first MALDI mass spectrometers started by two independent groups and was completed in 1990s (Dudley, 2014). MALDI IMS involves mixing of the molecules of interest from the sample with the matrix solution (usually organic acids) and co-crystallization of them together (Aichler and Walch, 2015). By applying pulsed UV-laser to the sample inside of MALDI ion source, matrix will absorb the energy and molecule will be desorbed and ionized into the gas phase (Figure 4). Depending on the choice of matrix type and technology, different analyte classes such as peptides, proteins, lipids, oligonucleotides and drugs can be investigated (Norris and Caprioli, 2013). The ion sources in MALDI IMS are combined with time of flight (TOF) chambers, where ions getting separated (Yalcin and de la Monte, 2015). Produced equal charged ions will have the same kinetic energy, but variation in ion velocities based on different masses. Therefore, by separating ions in high vacuum of TOF analyzer, smaller particles will move faster and reach the detector earlier than heavy particles, what will be determined in mass-to-charge ratio (m/z) (Wiley and McLaren, 1955). As a result the molecular maps with spatial distribution of m/z value of interest for each x and y coordinate will be generated. For further analysis obtained MALDI image can be compared to a histological stained microscope image (like hematoxylin and eosin stain (H&E), immunohistochemistry et al.) and co-registered to the mass spectrometric data (Aichler and Walch, 2015; Gessel et al., 2014; Wiley and McLaren, 1955).

Depending on the experimental purposes MALDI IMS analysis can involve a number of steps, however the critical step is a complex sample preparation on behalf of optimization of sensitivity and spatial resolution. Typically, sample preparation starting

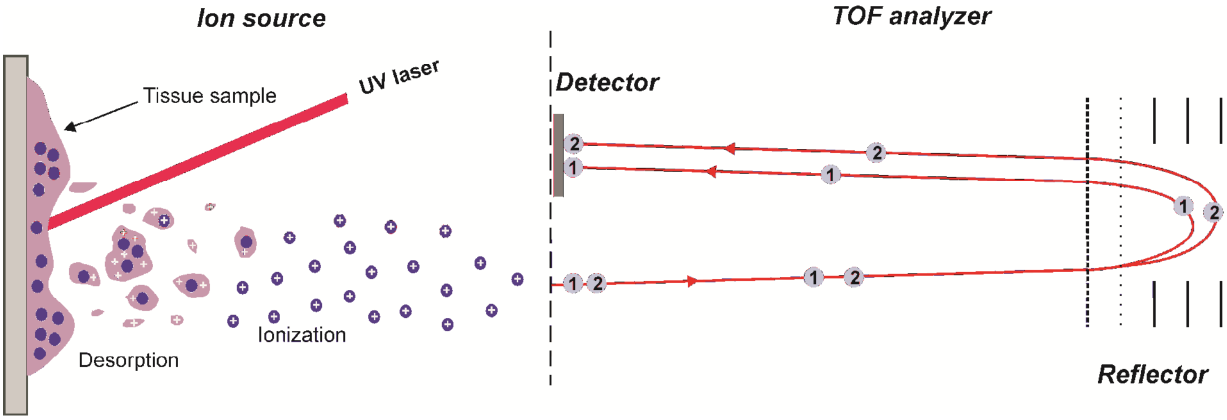


Figure 4. Principle of MALDI IMS. Adapted from Marvin *et al.* (Marvin et al., 2003).

with mounting thin tissue slice on conductive slide or sample homogenization, followed by centrifugation, extraction, washing or fixation steps (Dudley, 2014; Groseclose et al., 2008; Meding and Walch, 2013). Afterwards coming to the choice of matrix and heterogeneity of matrix crystallization and its application on a sample (Norris and Caprioli, 2013). Furthermore, two additional steps are necessary to be performed before the MALDI IMS measurement will take place (Yalcin and de la Monte, 2015). The first step is a calibration of the system to protein or peptide standards for specificity of obtained m/z ratio. The second one is to mark the sample with three teaching points for future coregistering with histological stained microscope image. The final step, after the MALDI IMS measurements were completed, is statistically correct interpretation of complex and extensive mass spectrometric data sets. It can be performed by standard control software provided by the inventors (e. g. Bruker Daltonik GmbH, Bremen, Germany) and by software programs introduced in recent years for processing MALDI IMS data (Norris and Caprioli, 2013).

By been able to detect tissue biomarkers, classify tumors, provide early diagnosis or prognosis, elucidate the pathways of pathogenesis and monitor therapy, MALDI IMS mass spectrometry has been applied for investigations of various diseases with the primary focus on cancer (Rodrigo et al., 2014). Over recent years, MALDI imaging has been used for investigation of gastrointestinal cancer, cancer of respiratory system, bladder, prostate, breast and ovaries (Alexandrov et al., 2013; Cheng et al., 2012; Chuang et al., 2012; Longuespée et al., 2012; Meding et al., 2012; Oezdemir et al., 2012; Rauser et al., 2010; Végvári and Döme, 2011). Further applications of MALD IMS in the field of gastroenterology and hepathology will be described in Chapter 4.2. Furthermore, the advantages and disadvantages of MALDI IMS mass spectrometry for biomedical applications are mentioned in the Table 3.

Table 3. Advantages and disadvantages of using MALDI IMS imaging mass spectrometry for biomedical applications

Advantages	**Disadvantages**
+ Fast + Label-free + Selective analysis of specific molecules + Analysis of samples with large size + Suitable for large number of samples + Possibility of direct identification of the analyte + Possibility of automatisation	- Complex and expensive sample preparation - Low reproducibility - Limits in spatial resolution - Long measurement time - Limited detectable mass range - Not suitable for *in vivo* application in the clinic

2. Aim of the work

The aim of this work was to develop new classification models for identification and classification of different types of liver cancer based on novel imaging techniques. Therefore, Raman spectroscopic imaging as a non-invasive, non-destructive and label-free method was used in order to develop classification models for liver cancer identification and prediction on both cellular and tissue levels. Furthermore, MALDI IMS mass spectrometry as a second method for analysis of specific biomarkers in malignant and non-malignant tissue regions in liver was applied.

The following aims were central in our research investigations:

1. To characterize and classify different types of liver cancer cells by Raman spectroscopic imaging.
2. To investigate metabolic changes during proliferation and apoptosis in HCC cells based on Raman spectra.
3. To test the reproducibility of Raman spectroscopic imaging as a diagnostic method.
4. To investigate molecular information of different hepatic lesions by Raman imaging aiming to classify and predict HCC.
5. To develop an algorithm for comparative analysis of main macromolecular components in liver cancer tissue sections.
6. To detect the borders of liver tumor in the hepatic specimens for surgical margin status application.
7. To apply MALDI IMS for identification of specific biomarkers associated with HCC.
8. To develop new classification model for diagnostics of HCC based on protein signatures detected by MALDI IMS.

3. Peer-reviewed publications

Publication 1:

Discrimination and Classification of Liver Cancer Cells and Proliferation States by Raman Spectroscopic Imaging

---

Tatiana Tolstik, Claudio Marquardt, Christian Matthäus, Norbert Bergner,
Christiane Bielecki, Christoph Krafft, Andreas Stallmach, Jürgen Popp

Published in:
Analyst, **2014**; 139(22):6036-43.
IF (2014): 4.107.

In this study the investigation of different liver cancer cell lines, i.e. HepG2 - well differentiated hepatocellular carcinoma cell line and SK-Hep1 - liver cancer cell line from a patient with adenocarcinoma, applying Raman imaging was performed. The identification and classification of different types of liver cancer cells (n = 100) by applying a support vector machine (SVM) classification algorithm allow to predict cell types with accuracy of 93 %. It was found out that the main difference between both cell lines can be attributed to Raman signatures of unsaturated fatty acids. Different cellular compartments could be distinguished by hierarchical cluster analysis (HCA) and analyzed separately. Therefore, separate use of spectra of the cytoplasmic lipids allows us to improve the classification to 96 %, compared to nucleus (accuracy of 91 %), cytoplasm (accuracy of 87 %) or the complete spectroscopic information of investigated cells (accuracy of 93 %). Furthermore, high expression of unsaturated fatty acids was detected in the proliferation phase of HCC cells which allowed us to classify and predict it vs. non-proliferation phase with sensitivity of 100 % and specificity of 98 %.

Analyst

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/c4an00211c

Discrimination and classification of liver cancer cells and proliferation states by Raman spectroscopic imaging

T. Tolstik,†*[ab] C. Marquardt,†[a] C. Matthäus,[b] N. Bergner,[b] C. Bielecki,[a] C. Krafft,[b] A. Stallmach[a] and J. Popp[bc]

Received 28th January 2014
Accepted 19th September 2014

DOI: 10.1039/c4an00211c

www.rsc.org/analyst

Discrimination of nodular lesions in cirrhotic liver is a challenge in the histopathologic diagnostics. For this reason, there is an urgent need for new detection methods to improve the accuracy of the diagnosis of liver cancer. Raman imaging allows to determine the spatial distribution of a variety of molecules in cells or tissue label-free and to correlate this molecular information with the morphological structures at the same sample location. This study reports investigations of two liver cancer cell lines, – HepG2 and SK-Hep1, – as well as HepG2 cells in different cellular growth phases using Raman micro-spectroscopic imaging. Spectral data of all cells were recorded as a color-coded image and subsequentially analyzed by hierarchical cluster and principal component analysis. A support vector machine-based classification algorithm reliably predicts previously unknown cancer cells and cell cycle phases. By including selectively the Raman spectra of the cytoplasmic lipids in the classifier, the accuracy has been improved. The main spectral differences that were found in the comparative analysis can be attributed to a higher expression of unsaturated fatty acids in the hepatocellular carcinoma cells and during the proliferation phase. This corresponds to the already examined *de novo* lipogenesis in cells of liver cancer.

1. Introduction

A range of optical techniques such as infrared, fluorescence, Raman and CARS spectroscopy have been widely used in recent years and proven to be reliable methods for cancer diagnostics.[1–9] Raman microscopic imaging is an innovative technology, which combines the spatial resolution of microscopic imaging with the highly specific spectroscopic information of classical Raman spectroscopy. It allows excitation of the vibrational and rotational states of molecules in biological samples in a label-free manner at submicron resolution. Hence, the biochemical composition of the sample can be examined at the subcellular level. Raman spectra of a biological sample are complex and result in a biochemical fingerprint, containing information about the chemical structure of proteins, nucleic acids, lipids and carbohydrates. Changes in these chemical structures in cells detected by Raman imaging can therefore help to differentiate and classify malignant tumors. For the evaluation of the spectral information, different multivariate unsupervised (*e.g.* clustering) and supervised (*e.g.* classification) methods have been employed.[1,10–32] Raman imaging has already successfully been applied for the identification of cell lines from the same and different origins,[1,10–17,28–31] stem cells,[18,19,32] the proliferative status of cells[20] and preliminary stages of cell apoptosis.[21–23]

The histopathological evaluation of biopsies is a crucial step to distinguish unclear nodular lesions in cirrhotic liver. With an occurrence rate of more than 85%, the hepatocellular carcinoma (HCC) is the most prevalent malignant primary liver tumor worldwide.[33] This second most lethal cancer with a 5 year survival rate of 8.9%[34] occurs with great frequency in Asia and Africa and is increasingly found in Europe. By an earlier diagnosis and an appropriate treatment the survival rate of patients with this malignant tumor can be improved. Despite a number of common staining techniques, morphological features have not proven to be clinically significant for HCC diagnosis.[35] Therefore, it is often difficult for the pathologist to diagnose the underlying disease just by cytohistologic criteria according to the International Consensus Group for Hepatocellular Neoplasia. Consequently, new detection methods are needed to reinforce the diagnosis of HCC.

In order to support the diagnosis of liver cancer, the aim of this research study was to characterize and differentiate two types of liver cancer cells by Raman spectroscopy and to prove

---

[a] *Department of Internal Medicine IV, Division of Gastroenterology, Hepatology and Infectious Diseases, Jena University Hospital, Erlanger Allee 101, 07747 Jena, Germany. E-mail: tatiana.tolstik@med.uni-jena.de; Fax: +49-03641-9324222; Tel: +49-03641-9324221*

[b] *Institute of Photonic Technology, Albert-Einstein-Straße 9, 07745 Jena, Germany. E-mail: juergen.popp@uni-jena.de; Fax: +49-03641-206099; Tel: +49-03641-2060*

[c] *Institute of Physical Chemistry and Abbe Center of Photonics, Friedrich-Schiller University Jena, Helmholtzweg 4, 07743 Jena, Germany. E-mail: juergen.popp@uni-jena.de; Fax: +49-03641-948302; Tel: +49-03641-948320*

† Both authors contributed equally.

its reproducibility. As an *in vitro* model system the HepG2 cell line (human cell line that was derived from liver tissue of a patient with well differentiated hepatocellular carcinoma) and the SK-Hep1 cell line (received from ascitic fluid of a patient with adenocarcinoma of the liver) were chosen for investigations by Raman imaging. Furthermore, we determined the Raman spectra of hepatocellular carcinoma cells (HepG2) in the exponential and plateau phase of cell growth. These two approaches were performed to detect Raman signatures and to generate classification models for two different types of liver cancer cells and also metabolic changes in liver cancer cells during a high cell division rate. The cell type and the proliferation behaviour are essential distinguishing features of malignant tumors. The results suggest that Raman spectroscopy is a particularly suitable method to easily identify molecular changes in liver cancer.

2. Materials and methods

2.1. Cell cultivation

HepG2 and SK-Hep1 cells were cultivated in RPMI 1640 liquid medium with 20 mM HEPES, stable glutamine (FG 1235; Biochrom AG, Germany), 10% fetal bovine serum (DE14-801F; Lonza, Belgium) and 1% penicillin–streptomycin (15140; Gibco®, Life Technologies GmbH, Germany). Both cell types were maintained in a 5% CO_2 incubator at 37 °C in cell culture flasks (658170; Greiner Bio-One GmbH, Germany). Every two days the medium was changed until approximately 50% confluence was reached. After that the cells were detached with trypsin–EDTA solution (L2143; Biochrom AG, Germany) and transferred onto CaF_2 slides at a concentration of 60 000 cells per mL. The Raman measurements were performed on CaF_2 slides in order to avoid autofluorescence background from regular glass slides.

2.2. Preparation of HepG2 and SK-Hep1 cells

HepG2 and SK-Hep1 cells were incubated for 24 hours, washed twice with phosphate buffered saline (PBS) (0689; Lonza, Belgium) and fixed with 10% neutral buffered formalin solution (HT501128; Sigma-Aldrich, USA) for 20 minutes. One more washing step was performed after the fixation process. Finally, the slides were stored in PBS buffer at +4 °C until the Raman measurements were done.

2.3. Preparation of exponential and plateau growth phase of HepG2 cells

To measure HepG2 cells with Raman spectroscopy in the exponential phase, the cells were cultivated in fresh medium for 24 h on CaF_2 slides. In order to achieve maximum cell proliferation, this cultivation process was performed until a monolayer culture with a confluence of 50% was reached. For the Raman measurement of HepG2 cells in the plateau phase, the monolayer culture was cultivated for 96 h until 100% confluence was obtained. Throughout the period of cell cultivation the medium was not exchanged, which leads to an inadequate nutrition of the cells and thus to a growth arrest (plateau phase). Afterwards, the samples were fixed with neutral buffered formalin solution for 20 minutes. Subsequently a washing step was performed and the samples were stored in PBS buffer at +4 °C.

2.4. Raman image acquisition

Raman spectra were acquired using a confocal Raman microscope (WITec, Ulm, Germany, Model CRM 2000) equipped with an air-cooled argon ion laser (LASOS Lasertechnik GmbH, Jena, Germany) with a laser power of *ca.* 5–10 mW at the sample (grating 600 g mm^{-1}, BLZ = 500 nm). An excitation wavelength of 488 nm was chose for our measurements because Raman scattering intensity is directly proportional to λ^{-4} (λ = laser wavelength), and therefore a blue laser with a shorter wavelength of 488 nm results in an increase in scattering intensity in compare with the longer wavelength. The raw Raman spectra acquired with 488 nm laser (Fig. 1) showed high signal to noise ratio with an acceptable fluorescent signal. The laser was coupled to a microscope by a single mode optical fiber. The Raman scattered light was detected by a back-illuminated deep-depletion CCD camera operating at −65 °C. The system was pre-calibrated to the 520.7 cm^{-1} spectral line of silicon and the laser light was focused on the sample with a 60 × NA=1.0 water immersion objective (Nikon NIR Apo, Tokyo, Japan). All samples were prepared in four repeats and the Raman measurements were performed in PBS solution. Raman spectral images of the cells were detected in the spectral region of 4100–200 cm^{-1} with a step size of 0.5 μm and an integration time for each spectrum of 1 second. Subsequently after the measurement the cells were stained with haematoxylin and eosin (HE) for morphological confirmation using a classical HE-staining protocol.

2.5. Data analysis

For the statistical analysis the complete spectral data set of all cells, which were recorded by the WITec software (WITec GmbH, Germany), were imported and calculated with the software packages CytoSpec (CytoSpec Inc., USA), OPUS (Bruker Optik GmbH, Germany) and MatLab (MathWorks Inc., USA) including the PLS_Toolbox (Eigenvector Research Inc., USA).

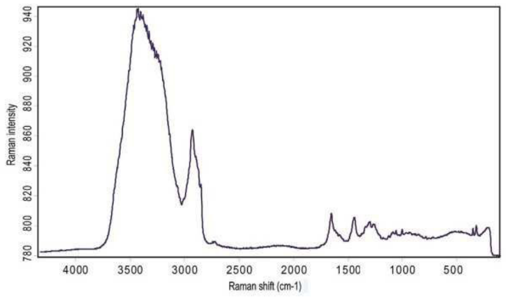


Fig. 1 Raw Raman spectra in the spectral region 4200–200 cm^{-1} of one HepG2 cell measured in water before any preprocessing steps.

For preprocessing, baseline correction and vector normalization in the spectral region of 3100–600 cm^{-1} were chosen,[36] which have been demonstrated to be appropriate and are routinously used for Raman data analysis.[12,37,38] A hierarchical cluster analysis (HCA), which was calculated with a Euclidean distance measure, clustered with the Ward's algorithm and filtered with the Generalized Least Squares Weighting, was applied to the complete data set of the cell lines, in order to split and identify new groups with a high level of similarity in the data (inter-individual cell variability). The relations between the data were illustrated in a dendrogram. For additional visualization of the data set, as a Raman image, the CH stretching intensities of the detected spectra at each measuring point were plotted. These color-coded Raman intensity maps were correlated with the underlying cytological compartments (cell organelles) and the differences in the data were displayed using the HCA. The HCA is grouping similar spectral information into different clusters, showing the spatial distribution of these clustered spectra corresponding to components of the cell (*e.g.* cell nucleus, cytoplasm, lipid storage organelles) as a Raman image. All spectra of each cell were grouped into five clusters including the spectra of nucleus, two different components of cytoplasm, cytoplasmic lipids and background spectra (intra-individual cell variability). The surrounding cluster of the cell background was removed from the data set. An average spectrum of each cell cluster and a difference spectrum were calculated with the OPUS software. Finally, the averaged data were imported into MatLab.

Principal component analysis (PCA) was applied to normalized and mean-centred Raman spectra. More than 99% of the data variance were represented by the first 20 principal components (PCs) and only noise was found in the higher PCs. In order to recognize a predictive pattern in the spectra, the first 20 PCs of a training data set were used to develop a supervised learning model using the support vector machine (SVM) algorithm, which is capable to classify future samples of an unknown validation set. Besides other classification algorithms, the SVM have already been successfully applied in various Raman imaging studies.[12,25,28,38] To estimate the performance of this predictive classification model a 10-fold cross-validation was performed.

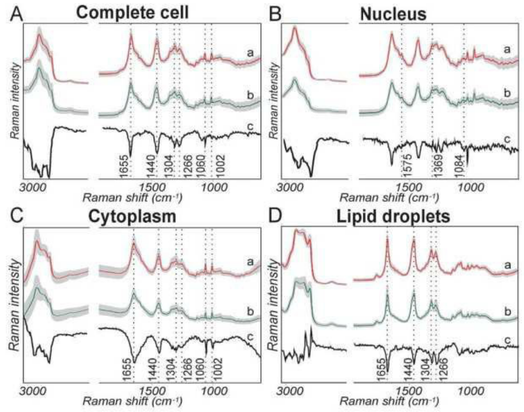


Fig. 2 Average spectra in the spectral region 3100–2800 cm^{-1} (×3) and 1800–600 cm^{-1} (×5) of (a) HepG2 cells, (b) SK-Hep1 cells, and (c) the corresponding difference spectra for (A) complete cell, (B) nucleus, (C) cytoplasm, and (D) lipid droplets. Distinguishing Raman bands (in cm^{-1}) are indicated by dashed lines. Shaded areas of the spectra a and b represents the standard deviations.

3. Results

3.1. Unsupervised analysis and classification of HepG2 and SK-Hep1 cells

By analysis of the Raman image data sets, acquired from 100 cells of HepG2 and SK-Hep1 cell lines, spectral differences based on the cytology have been identified. The average spectra for each cell line and each cellular compartment in the spectral region of 3100–500 cm^{-1} are presented in Fig. 2. The Raman spectra show the typical characteristics of cells, such as CH stretching intensities (3020–2800 cm^{-1}), amide I band (1680–1620 cm^{-1}), CH_2 deformation band (1440 cm^{-1}) and the phenylalanine band (1002 cm^{-1}). By subtraction of the average spectra of the HepG2 and SK-Hep1 cells, the spectral variations, shown as a difference spectrum, were obtained (Fig. 2c). The difference spectrum exhibits intensities in the range of 2900–2850 cm^{-1} and at 1655, 1440, 1304, 1266, 1060 cm^{-1}.

3.1.1. Hierarchical cluster analysis of inter- and intra-individual cell variabilities.

For the detection of the inter-individual variability of the two different cell lines a HCA of all HepG2 ($n = 48$) and SK-Hep1 ($n = 52$) cells was performed. This way, a clear separation of HepG2 and SK-Hep1 cells was achieved based on the cell-type specific Raman pattern (100% accuracy). The plotted dendrogram in Fig. 3 displays the distinct subdivision according to the cell line.

For the detection of the intra-individual variability each single cell were subdivided by HCA in five clusters corresponding to the different cell compartments, which were confirmed by HE stained microscope images and Raman maps

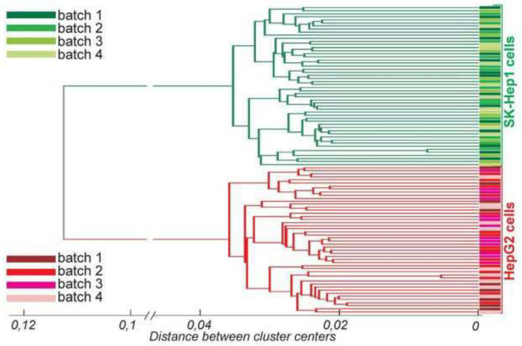


Fig. 3 Dendrogram from HCA of Raman spectra showing the clustering of HepG2 ($n = 48$) and SK-Hep1 ($n = 52$) cells with assigned subclusters. The batches of the cell lines are marked in different colors.

of CH-stretching region (Fig. 4A–C). In accordance with the cell nucleus the average Raman spectra showed the typical bands of the nucleic acids at 1575, 1369, 1084 and 782 cm^{-1}, whereas in the average spectra of the cytoplasm spectral pattern of proteins and lipids without bands of nucleic acids were observed (Fig. 4D and E). The data regions of the lipid rich cell components, confirmed by the typical Raman spectra of lipids, demonstrated intensive bands at 3008, 2898, 2854, 1655, 1440, 1304, 1266, 1078, 1059, 968 and 717 cm^{-1} (Fig. 4F). The obtained clusters were used to extract the spectral data of each single cell compartment and to implement them in the SVM-based classification algorithm described below.

3.1.2. Principal component analysis. Using PCA the wide variety of statistical variables of the extensive spectral data set was converted into PCs. The bands at 2900, 2898, 2854, 1660, 1655, 1440, 1304, 1267, 1266, 1060 cm^{-1} were identified in the calculated PCs (Fig. 5). These bands resemble those bands of the difference spectrum (Fig. 2B) and are also similar to the bands of pure lipids spectra (Fig. 4F). Almost the half (48, 76%) of the variance in this data set was corresponded to PC1 and can be assigned to lipids (2848, 2854, 1655, 1440, 1304, 1266, and 1060 cm^{-1}).

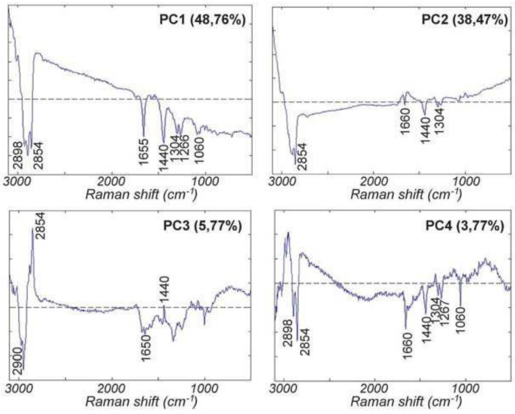


Fig. 5 (A) Loadings for the first four principal components of the PCA model of HepG2 and SK-Hep1 cell lines in a spectral region 3100–600 cm^{-1}.

3.1.3. Support vector machine classification model. A SVM algorithm was trained on the averaged spectra of 60 cells of HepG2 ($n = 30$) and SK-Hep1 ($n = 30$) to define a Raman pattern that predicts the cell type. For the recognition algorithm the first 20 PCs were used of which the first four PCs captured the main variance (accounting for 96.8%). The PCs higher than 20, which mainly contained noise and had less than 1% of the total variance, were not included in the classification algorithm. A performed 10-fold cross-validation, which was calculated by using subsets of the trainings set, estimates the predictive performance of this classification algorithm with 89%. The cross-validation classified not more than one HepG2 cell incorrectly as SK-Hep1 cell and a maximum of seven SK-Hep1 cells were misclassified as HepG2. By the following obtained classification model previously unknown cells of the two cell lines from a validation set (HepG2: $n = 18$; SK-Hep1: $n = 22$) were identified correctly with a prediction accuracy of 93% (sensitivity of 97%, specificity of 89%).

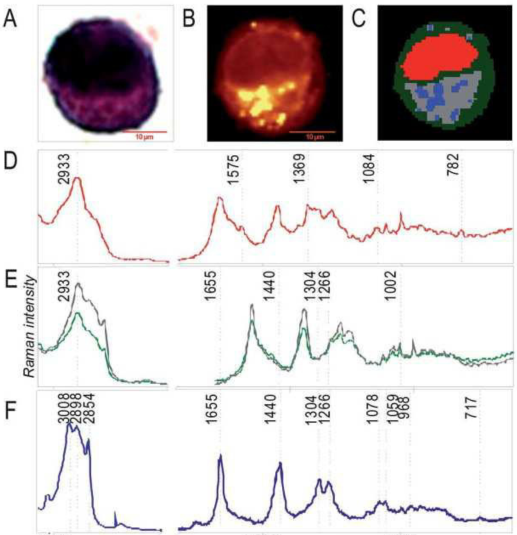
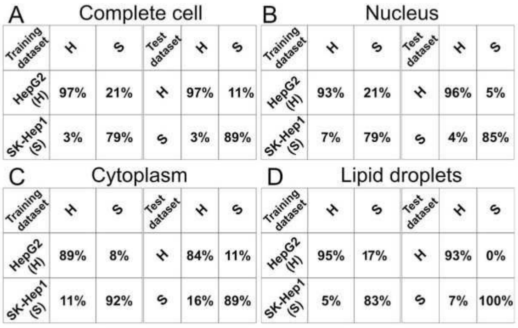


Fig. 4 (A) HE stained microscopy image, (B) Raman image of the CH-stretching region. (C) HCA map with five clusters of the same HepG2 cell, including the cluster for nucleus (red), cytoplasm (grey and green) and lipid droplets (blue). (D–F) Corresponding average spectra to the cluster of (D) nucleus, (E) cytoplasm, and (F) lipid droplets in the spectral region 3100–2800 cm^{-1} (×1) and 1800–600 cm^{-1} (×5).

In addition to the classification model of complete cells, we elaborated also the difference spectra and classification algorithm by using the spectral information of different cell compartments. For this purpose, a discriminant analysis was performed and the data were imported in the SVM-algorithm. Consequently, for each cell compartment, such as the cell nucleus, cytoplasm and the lipid storage organelles, separate classification models were generated. By using both specific data sets as two independent classifiers an accuracy of 91% and 87% for the data of cell nucleus and cytoplasm were obtained, respectively. The spectral information of lipid droplets, confirmed by the typical Raman spectra of lipids, were isolated by HCA and also used to classify HepG2 and SK-Hep1. This classifier showed a very high sensitivity of 93%, specificity of 100% and accuracy of 96%. The details of the prediction for individual classification model are shown in the confusion table of Fig. 6.

3.1.4. Test-retest reliability of the sample measurements. To verify the test-retest reliability of the generated classification algorithm two different data sets of the HepG2 and SK-Hep1 cell lines were acquired under the same conditions, but at different points in time. Data sets were collected from Raman



A Complete cell

Training dataset	H	S	Test dataset	H	S
HepG2 (H)	97%	21%	H	97%	11%
SK-Hep1 (S)	3%	79%	S	3%	89%

B Nucleus

Training dataset	H	S	Test dataset	H	S
HepG2 (H)	93%	21%	H	96%	5%
SK-Hep1 (S)	7%	79%	S	4%	85%

C Cytoplasm

Training dataset	H	S	Test dataset	H	S
HepG2 (H)	89%	8%	H	84%	11%
SK-Hep1 (S)	11%	92%	S	16%	89%

D Lipid droplets

Training dataset	H	S	Test dataset	H	S
HepG2 (H)	95%	17%	H	93%	0%
SK-Hep1 (S)	5%	83%	S	7%	100%

Fig. 6 Confusion tables of the SVM classification models of HepG2 (H) and SK-Hep1 (S) cells. Misclassifications (in %) of (A) complete cell, (B) nucleus, (C) cytoplasm, and (D) lipid droplets in training and test dataset are shown.

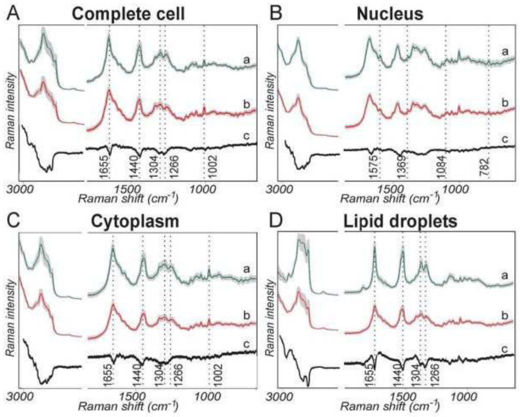


Fig. 7 Average Raman spectra of (a) proliferating and (b) non-proliferating HepG2 cells, and (c) the corresponding difference spectra of (A) complete cell, (B) nucleus, (C) cytoplasm, and (D) lipid droplets. Shaded areas of the spectra a and b represents the standard deviations.

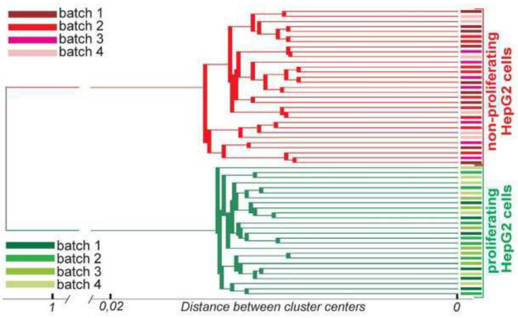


Fig. 8 Hierarchical clustering of proliferating HepG2 cells in the exponential phase ($n = 26$) and the non-proliferating HepG2 cells in the plateau phase ($n = 30$) demonstrated as a dendrogram with assigned sub-clusters. The batches of the cells are marked in different colors.

measurements of three independent samples of HepG2 and SK-Hep1 cells from different batches and with a time lag of 3 months. The second data set was used to train the classification model. Therefore, the prediction accuracy of the classification model was first estimated by performing a 10-fold cross-validation of the 100 cells. The cross-validation calculated a sensitivity of 94% and a specificity of 87% for the predictive accuracy. However, the trained classification model based on the first data set applied to the independent second data set resulted then in a sensitivity of 82% and a specificity of 75%.

3.2. Unsupervised analysis and classification of different proliferation performances of HepG2 cells

The difference spectrum of the proliferating (exponential phase) and non-proliferating (plateau phase) HepG2 cells indicated higher intensities within the spectral region between 3020–2800 cm^{-1} and at 1655, 1440, 1304, 1266 and 1002 cm^{-1} of the proliferating cells (Fig. 7A).

3.2.1. Hierarchical cluster analysis. To detect also the inter-individual variability in case of HepG2 cells in the exponential ($n = 26$) and plateau ($n = 31$) growth stages HCA was performed and plotted as a dendrogram. The tree structure in the dendrogram (Fig. 8) shows a perfect discrimination of the proliferating and non-proliferating HepG2 cells in two separated clusters. Thereby, all HepG2 cells were clustered according to their growth phase and not one cell was misclassified (100% accuracy).

Each single HepG2 cell in different proliferating condition was as well clustered in intra-individual components of the cell. The extracted spectral information of the correspondent nucleus, cytoplasm and lipid droplet regions was used for classification (Fig. 7).

3.2.2. Principal component analysis. Analyzing PCs of this data set the bands 2888, 2851, 1655, 1440, 1304, 1266, and 1084 cm^{-1} seen in PC1 includes more than 93% of variance and are equivalent to lipid spectra (Fig. 9A). The positive bands that were detected in PC2 (Fig. 9B) corresponding to DNA (1577, 1096 and 790 cm^{-1}) and proteins (2935, 1676, 1340, 1250 and 1005 cm^{-1}), while the negative bands can be assigned to lipids (2848 and 1434 cm^{-1}).

3.2.3. Support vector machine classification model. In order to classify the HepG2 cells in different cellular growth stage, a training and validation data set of cells in the exponential and plateau phases were calculated by the SVM-based classification model. The performance of the classification model to predict the unknown proliferation phase of HepG2 cells exhibited a sensitivity of 100% and specificity of 98%. As mentioned above for the classification of HepG2 and SK-Hep1, spectra of proliferating and non-proliferating HepG2 cell compartments were classified separately. Over all accuracies of the classification models the classifier based on the Raman spectra of lipid droplets organelles showed the best result. The details of misclassified proliferating and non-proliferating HepG2 cells are demonstrated in the confusion table in Fig. 10.

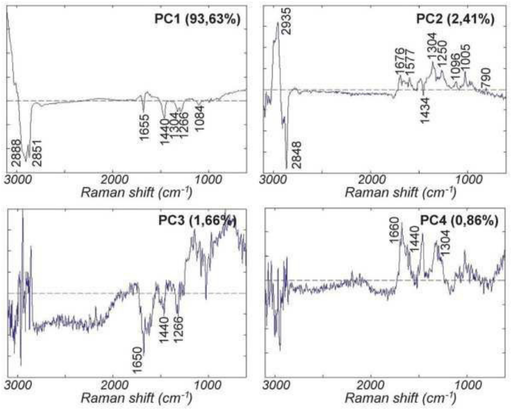
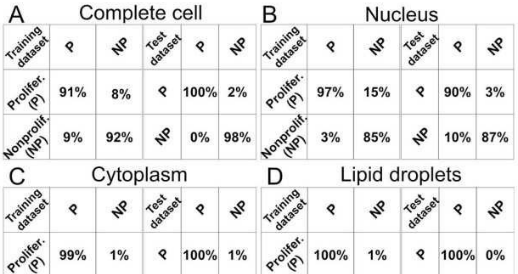


Fig. 9 First four PC from PCA model performed on proliferating (exponential phase) and non-proliferating (plateau phase) HepG2 cells.

4. Discussion

To prove the concept that Raman imaging technique is suitable for the detection of malignant cells in liver lesions the morphological and proliferating features of cancer cells were investigated and classified by this vibrational spectroscopic method.

In previous studies it has already been demonstrated that it is possible to detect differences in normal and malignant hepatocytes based on single point Raman spectra.[13,38] Therefore, the focus of this study was set on the differentiation and classification of different tumor cells. In contrast to previous studies, the differentiation was based on spectral information collected from whole cell, rather than from single point spectra with a sampling volume of about 1 μm^3. A liver cancer cell line of the hepatocellular carcinoma (HepG2), as the most common primary liver cancer, and an adenocarcinoma (SK-Hep1), that is found as primary liver cancer in the liver, were measured with Raman spectroscopy. The obtained Raman spectra were processed with a multivariate analysis method and implemented into a reliable prediction model. Thus, to the best of our knowledge this is the first study that is using vibrational spectroscopy as a classifying tool for different types of liver cancer cells and their proliferation states.



A Complete cell

Training dataset	P	NP	Test dataset	P	NP
Prolifer. (P)	91%	8%	P	100%	2%
Nonprolif. (NP)	9%	92%	NP	0%	98%

B Nucleus

Training dataset	P	NP	Test dataset	P	NP
Prolifer. (P)	97%	15%	P	90%	3%
Nonprolif. (NP)	3%	85%	NP	10%	87%

C Cytoplasm

Training dataset	P	NP	Test dataset	P	NP
Prolifer. (P)	99%	1%	P	100%	1%
Nonprolif. (NP)	1%	99%	NP	0%	99%

D Lipid droplets

Training dataset	P	NP	Test dataset	P	NP
Prolifer. (P)	100%	1%	P	100%	0%
Nonprolif. (NP)	0%	99%	NP	0%	100%

Fig. 10 Confusion tables of the SVM classification models of proliferating (P) and non-proliferating (NP) HepG2 cells. Misclassifications (in %) of (A) complete cell, (B) nucleus, (C) cytoplasm, and (D) lipid droplets in training and test dataset are shown.

Like a molecular fingerprint the Raman spectra exhibit spectral differences of cells and to some extent molecular contents of various cellular components, such as nucleic acids, proteins, carbohydrates and lipids. While the specificity of Raman spectroscopy is very high, the signal intensity and thus the level of sensitivity of this technique is quite low, which can cause a reduced detection of molecules with low concentration. Other limitations of this technique can arise from an unwanted background signal originated from the fluorescence of the sample or contamination.

In our investigations the difference spectrum of the observed averaged spectra and the multivariate discriminant analysis of HepG2 and SK-Hep1 data sets were able to extract the subtle spectral differences between these two cell lines with pronounced intensities in the C–H stretching region of 2900–2850 cm^{-1} and a characteristic shape in the fingerprint region between 1800–600 cm^{-1}. The results represent significant spectral differences in the peak intensities at 2900–2850, 1655, 1440, 1304, 1266 and 1060 cm^{-1}. This recognition pattern applied in a multivariate – HCA allows already a precise separation of the HepG2 and SK-Hep1 cell line in two major clusters without any misclassification. Furthermore, by implementing the same spectral pattern into the above-mentioned classification algorithm the previously unknown cell identity can be predicted with an accuracy of 93%. In the discriminant analysis the average spectra of both cell lines differ significantly in the specific spectral characteristics of lipids. Due to this fact, the prediction accuracy of the classification model was improved to 96% by the sole use of the spectral pattern of the cytoplasmic lipids. Thus, it was possible to show that a majority of spectral information, which characterizes and distinguishes different liver cancer cells, is located in lipids. These detected spectral differences can be assigned exactly to the specific wavenumbers of unsaturated fatty acids and were verified by a contrasting juxtaposition of the averaged spectra according to the lipid clusters to the reference spectra of unsaturated fatty acids (Fig. 11a). Hence, the quantity of unsaturated acids is increased in the HepG2 cell line.

By inspecting liver cancer cells with Raman microscopic imaging, the results confirm that a lot of molecular differences are hidden in lipids. Nowadays, it is believed that aberrant lipid biosynthesis is involved in the hepatocarcinogenesis. Already in 2005, the mRNA expression of lipogenic enzymes, which are involved in the production of fatty acids, was investigated in 10 human HCC tissue samples and were compared with the surrounding non-cancerous liver tissue. An elevated expression of mRNA for fatty acid synthase (FASN), acetyl-CoA carboxylase (ACAC) and ATP citrate lyase (ACLY) in the HCC tissue was described.[39] In a recent large-scale study the expressions of enzymes that regulate lipogenesis were evaluated in HCC

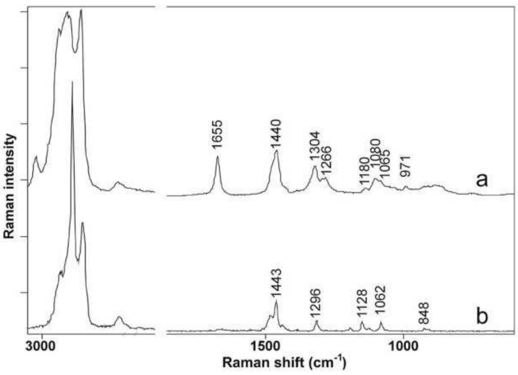


Fig. 11 Reference Raman spectra of (a) unsaturated fatty acid (oleic acid) and (b) saturated fatty acid (stearic acid) acquired with WITec Raman spectrometer.

tissues and corresponding surrounding non-tumor liver tissues of 68 patients as well as in eight normal (disease-free) human liver tissue of healthy individuals. It was shown by this study that all relevant lipogenic enzymes, which are involved in the fatty acid synthesis (FASN, ACAC, ACLY, ME, SCD1), the cholesterol biosynthesis (SREBP2, HMGCR, MVK, SQS) and their upstream inductors (chREBP, SREBP1, LXR-β) were progressively upregulated from non-tumorous liver tissue toward the HCC. Analogously to the upregulation of the lipogenic enzymes and their upstream inductors also the quantity of the chemical end products of the respective lipid synthesis (fatty acids, triglycerides, cholesterol) increased. Furthermore, the study demonstrated that a higher expression of these enzymes correlate with an activation of a lipogenic signaling pathway (AKT–mTORC1–RPS6 pathway) and the clinical aggressiveness, characterized by shorter (<3 years) and longer (>3 years) survival of the patients following partial liver resection.[40] In another study of three HCC cell lines, it has been observed that the FASN protein expression and basal activity levels were higher in HepG2 cells than in the other two HCC cell lines. The inhibition of the FASN activity with C75 induced a cell cycle arrest in all three HCC cell lines.[41]

As set out in the study of Gao *et al.*[41] in our comparative analysis also a higher content of fatty acids in the HepG2 cells were detected by the Raman imaging technology. Additionally, the amount of fatty acids in proliferating HepG2 cells obtained by Raman imaging in the exponential phase was higher than in the plateau phase, in which the cells are usually not proliferating. These findings are also consistent with data of Gao *et al.*,[41] which indicate that an inhibition of the fatty acid synthesis causes a cell cycle arrest. By performing a classification using the spectral pattern of HepG2 cells in the exponential and plateau phase the state of cell proliferation can be predicted with almost 100% accuracy (sensitivity of 100%, specificity of 98%).

In a previous study it was already possible to detect differences in normal and malignant hepatocytes based on single point Raman spectra.[13,42] Furthermore it was demonstrated that Raman spectroscopy is able to distinguish a metastatic from a non-metastatic cell line and to associate the detected spectral differences, which were assigned to a higher unsaturated fatty acid content, to the metastatic ability of the two isogenic cancer cell lines.[17]

5. Conclusions

In summary, this study has clearly demonstrated that Raman imaging spectroscopy in combination with multivariate data analysis is able to detect the cell-type specific molecular variations and the variation in cell proliferation of liver cancer cells. Furthermore, by using a support vector machine-based classification algorithm it was even possible to predict with a high accuracy the unknown identity and proliferation behaviour of liver cancer cells and these results have proven to be reproducible. The majority of the detected spectral differences were attributed to a larger quantity of unsaturated fatty acids in the HCC cells and during the proliferation phase. The cell type and the proliferation behaviour are essential distinguishing features of malignant tumors and Raman spectroscopy is a particularly suitable label-free method to identify easily this *de novo* lipogenesis in liver cancer.

Acknowledgements

We are grateful to co-workers form gastroenterology laboratory for support, Thomas Bocklitz for critical discussions and Marita Vetterlein for the excellent technical assistance. Financial support from the European Union *via* the "Europäischer Fonds für Regionale Entwicklung (EFRE)" and the "Thüringer Ministerium für Bildung, Wissenschaft und Kultur (TMBWK)" is highly acknowledged.

Notes and references

1 A. T. Harris, M. Garg, X. B. Yang, S. E. Fisher, J. Kirkham, D. A. Smith, D. P. Martin-Hirsch and A. S. High, *Head Neck Oncol.*, 2009, **1**, 38.
2 C. Krafft, B. Dietzek and J. Popp, *Analyst*, 2009, **134**, 1046–1057.
3 T. Upile, W. Jerjes, H. J. C. M. Sterenborg, A. K. El-Naggar, A. Sandison, M. J. H. Witjes, M. A. Biel, I. Bigio, B. J. F. Wong, A. Gillenwater, A. J. MacRobert, D. J. Robinson, C. S. Betz, H. Stepp, L. Bolotine, G. McKenzie, C. A. Mosse, H. Barr, Z. Chen, K. Berg, A. K. D'Cruz, N. Stone, C. Kendall, S. Fisher, A. Leunig, M. Olivo, R. Richards-Kortum, K. C. Soo, V. Bagnato, L.-P. Choo-Smith, K. Svanberg, I. B. Tan, B. C. Wilson, H. Wolfsen, A. G. Yodh and C. Hopper, *Head Neck Oncol.*, 2009, **1**, 25.
4 C. Krafft, B. Dietzek, M. Schmitt and J. Popp, *J. Biomed. Opt.*, 2012, **17**, 040801.
5 P. Matousek and N. Stone, *J. Biophotonics*, 2013, **6**, 7–19.
6 D. I. Ellis, D. P. Cowcher, L. Ashton, S. O'Hagan and R. Goodacre, *Analyst*, 2013, **138**, 3871–3884.

7 P. Chen, A. Shen, X. Zhou and J. Hu, *Anal. Methods*, 2011, **3**, 1257–1269.

8 A. Shen, B. Zhang, J. Ping, W. Xie, P. Donfack, S.-J. Baek, X. Zhou, H. Wang, A. Materny and J. Hu, *J. Raman Spectrosc.*, 2009, **40**, 550–555.

9 M. S. Bergholt, W. Zheng, K. Lin, K. Y. Ho, M. Teh, K. G. Yeoh, J. B. Y. So and Z. Huang, *Analyst*, 2010, **135**, 3162–3168.

10 M. Harz, M. Kiehntopf, S. Stöckel, P. Rösch, T. Deufel and J. Popp, *Analyst*, 2008, **133**, 1416–1423.

11 T. Bocklitz, M. Putsche, C. Stüber, J. Käs, A. Niendorf, P. Rösch and J. Popp, *J. Raman Spectrosc.*, 2009, **40**, 1759–1765.

12 U. Neugebauer, J. H. Clement, T. Bocklitz, C. Krafft and J. Popp, *J. Biophotonics*, 2010, **3**, 579–587.

13 S. R. Hawi, W. B. Campbell, A. Kajdacsy-Balla, R. Murphy, F. Adar and K. Nithipatikom, *Cancer Lett.*, 1996, **110**, 35–40.

14 C. Yu, E. Gestl, K. Eckert, D. Allara and J. Irudayaraj, *Cancer Detect. Prev.*, 2006, **30**, 515–522.

15 P. R. T. Jess, D. D. W. Smith, M. Mazilu, K. Dholakia, A. C. Riches and C. S. Herrington, *Int. J. Cancer*, 2007, **121**, 2723–2728.

16 Y. Oshima, H. Shinzawa, T. Takenaka, C. Furihata and H. Sato, *J. Biomed. Opt.*, 2010, **15**, 017009.

17 M. Hedegaard, C. Krafft, H. J. Ditzel, L. E. Johansen, S. Hassing and J. Popp, *Anal. Chem.*, 2010, **82**, 2797–2802.

18 J. W. Chan, D. K. Lieu, T. Huser and R. A. Li, *Anal. Chem.*, 2009, **81**, 1324–1331.

19 A. Ghita, F. C. Pascut, M. Mather, V. Sottile and I. Notingher, *Anal. Chem.*, 2012, **84**, 3155–3162.

20 K. W. Short, S. Carpenter, J. P. Freyer and J. R. Mourant, *Biophys. J.*, 2005, **88**, 4274–4288.

21 S. Verrier, I. Notingher, J. M. Polak and L. L. Hench, *Biopolymers*, 2004, **74**, 157–162.

22 C. Krafft, T. Knetschke, R. H. W. Funk and R. Salzer, *Anal. Chem.*, 2006, **78**, 4424–4429.

23 M. R. Guarracino, P. Xanthopoulos, G. Pyrgiotakis, V. Tomaino, B. M. Moudgil and P. M. Pardalos, *Artif. Intell. Med.*, 2011, **53**, 119–125.

24 N. Bergner, C. Krafft, K. D. Geiger, M. Kirsch, G. Schackert and J. Popp, *Anal. Bioanal. Chem.*, 2012, **403**, 719–725.

25 N. Bergner, T. Bocklitz, B. F. M. Romeike, R. Reichart, R. Kalff, C. Krafft and J. Popp, *Chemom. Intell. Lab. Syst.*, 2012, **117**, 224–232.

26 C. Krafft, M. A. Diderhoshan, P. Recknagel, M. Miljkovic, M. Bauer and J. Popp, *Vib. Spectrosc.*, 2011, **55**, 90–100.

27 C. Beleites, U. Neugebauer, T. Bocklitz, C. Krafft and J. Popp, *Anal. Chim. Acta*, 2013, **760**, 25–33.

28 U. Neugebauer, T. Bocklitz, J. H. Clement, C. Krafft and J. Popp, *Analyst*, 2010, **135**, 3178–3182.

29 S. Dochow, C. Krafft, U. Neugebauer, T. Bocklitz, T. Henkel, G. Mayer, J. Albert and J. Popp, *Lab Chip*, 2011, **11**, 1484–1490.

30 C. Krafft, B. Belay, N. Bergner, B. F. M. Romeike, R. Reichart, R. Kalff and J. Popp, *Analyst*, 2012, **137**, 5533–5537.

31 J. W. Chan, *J. Biophotonics*, 2013, **6**, 36–48.

32 J. W. Chan and D. K. Lieu, *J. Biophotonics*, 2009, **2**, 656–668.

33 D. M. Parkin, F. Bray, J. Ferlay and P. Pisani, *Ca-Cancer J. Clin.*, 2005, **55**, 74–108.

34 P. A. Farazi and R. A. DePinho, *Nat. Rev. Cancer*, 2006, **6**, 674–687.

35 M. Roncalli, Y. N. Park and L. Di Tommaso, *Dig. Liver Dis.*, 2010, **42**(suppl. 3), S228–S234.

36 P. Lasch, *Chemom. Intell. Lab. Syst.*, 2012, **117**, 100–114.

37 A. I. Mazur, J. L. Monahan, M. Miljković, N. Laver, M. Diem and B. Bird, *J. Biophotonics*, 2013, **6**, 101–109.

38 C. Bielecki, T. W. Bocklitz, M. Schmitt, C. Krafft, C. Marquardt, A. Gharbi, T. Knösel, A. Stallmach and J. Popp, *J. Biomed. Opt.*, 2012, **17**, 076030.

39 N. Yahagi, H. Shimano, K. Hasegawa, K. Ohashi, T. Matsuzaka, Y. Najima, M. Sekiya, S. Tomita, H. Okazaki, Y. Tamura, Y. Iizuka, K. Ohashi, R. Nagai, S. Ishibashi, T. Kadowaki, M. Makuuchi, S. Ohnishi, J. Osuga and N. Yamada, *Eur. J. Cancer*, 2005, **41**, 1316–1322.

40 D. F. Calvisi, C. Wang, C. Ho, S. Ladu, S. A. Lee, S. Mattu, G. Destefanis, S. Delogu, A. Zimmermann, J. Ericsson, S. Brozzetti, T. Staniscia, X. Chen, F. Dombrowski and M. Evert, *Gastroenterology*, 2011, **140**, 1071–1083.

41 Y. Gao, L.-P. Lin, C.-H. Zhu, Y. Chen, Y.-T. Hou and J. Ding, *Cancer Biol. Ther.*, 2006, **5**, 978–985.

42 J. Guo, B. Du, M. Qian, W. Cai, Z. Wang and Z. Sun, *Chin. Opt. Lett.*, 2009, **7**, 60–63.

Publication 2:

Classification and Prediction of HCC Tissues by Raman Imaging with Identification of Fatty Acids as Potential Lipid Biomarkers

---

Tatiana Tolstik, Claudio Marquardt, Claudia Beleites, Christian Matthäus,
Christiane Bielecki, Martin Bürger, Christoph Krafft, Olaf Dirsch, Utz Settmacher,
Jürgen Popp, Andreas Stallmach

Published in:

Journal of Cancer Research and Clinical Oncology, **2015 Mar;** 141(3):407-18.

IF (2014): 3.081.

Based on our previous study, Raman imaging was applied for tissue investigations in order to differentiate, classify and predict malignant and non-malignant regions of the liver. Thus, different components in a tissue such as lipids, proteins, collagen and cholesterol ester by N-finder algorithm (N-FINDER) were detected by Raman imaging. After analyzing their distribution in the samples the detection of the main components of HCC and fibrotic tissue regions was possible. The predefined malignant and non-malignant tissue regions were used for a comparative analysis and calculated Random forest (RF) classifier with a sensitivity of 76 % and specificity of 93 % to predict HCC. The most significant spectral differences were identified as fatty acids. Furthermore, three of the most significant bands were attributed to palmitic acid. Subsequently, we applied the developed classification algorithm to tumor margins in order to verify the predictive performance of the classifier. Predicted borders of the tumor were correlated with H&E stained tissue regions and showed high potential of Raman spectroscopic imaging for surgical margin status detection.

J Cancer Res Clin Oncol
DOI 10.1007/s00432-014-1818-9

ORIGINAL ARTICLE – CANCER RESEARCH

Classification and prediction of HCC tissues by Raman imaging with identification of fatty acids as potential lipid biomarkers

T. Tolstik · C. Marquardt · C. Beleites · C. Matthäus · C. Bielecki · M. Bürger · C. Krafft · O. Dirsch · U. Settmacher · J. Popp · A. Stallmach

Received: 25 August 2014 / Accepted: 27 August 2014

Abstract

Purpose Patients with hepatocellular carcinoma (HCC) can only be treated curatively at early stages and then have a favorable prognosis of this often fatal disease. For this reason, an early detection and diagnostic confirmation are crucial. Raman imaging spectroscopy is a promising technology for high-resolution visualization of the spatial distribution of molecular composition in tissue sections. The aim of this study was to investigate molecular information of liver tissue by Raman imaging for classification and diagnostic prediction.

Methods Unstained cryosections of human hepatic tissues (23 patients) were measured by Raman spectroscope in the regions of HCC ($n = 12$) and fibrosis ($n = 17$). The acquired data set was used to generate a random forest classification model with 101 iterations of sevenfold cross-validation. The models obtained during cross-validation were also used to predict regions of tumor margin ($n = 8$) aside from independent testing.

Results Raman spectra differed between malignant and non-malignant tissue regions. Based on these spectral data, a random forest classification model calculated a prediction accuracy of 86 % (76 % sensitivity and 93 % specificity). The ten most important variables were identified at 2895, 2856, 1439, 1298, 1080, 1063, 1023, 937, 920, and 719 cm^{-1}.

Conclusions In this study, Raman imaging spectroscopy was applied successfully for liver tissue to differentiate, classify, and predict with high accuracy malignant and non-malignant tissue regions. Furthermore, the most important differences were identified as the Raman signature of fatty acids. The demonstrated results highlight the enormous potential which vibrational spectroscopy techniques have for the future diagnostics and prognosis estimation of HCC.

Keywords Raman spectroscopic imaging · Hepatocellular carcinoma · Fibrosis · Liver tissue · Random forest classification model · N-FINDR

Introduction

The dramatic increase in the incidence of hepatocellular carcinoma (HCC) causes worldwide 692,000 deaths per year (Iavarone and Colombo 2013). Surveillance strategies and the early diagnosis of the tumor are essential for curative treatment option. Histopathological diagnosis of HCC is based essentially on the cytohistologic criteria according to the International Consensus Group for Hepatocellular

---

T. Tolstik · C. Marquardt · C. Bielecki · M. Bürger · A. Stallmach (✉)
Division of Gastroenterology, Hepatology and Infectious Diseases, Department of Internal Medicine IV, Jena University Hospital, Erlanger Allee 101, 07747 Jena, Germany
e-mail: andreas.stallmach@med.uni-jena.de

T. Tolstik · C. Beleites · C. Matthäus · C. Krafft · J. Popp (✉)
Leibniz Institute of Photonic Technology, Albert-Einstein-Straße 9, 07745 Jena, Germany
e-mail: juergen.popp@ipht-jena.de

O. Dirsch
Institute of Pathology, Friedrich-Schiller-University, Ziegelmühlenweg 1, 07743 Jena, Germany

U. Settmacher
Department of Visceral Surgery, Jena University Hospital, Erlanger Allee 101, 07747 Jena, Germany

J. Popp
Institute of Physical Chemistry and Abbe Center of Photonics, Friedrich-Schiller University Jena, Helmholtzweg 4, 07743 Jena, Germany

Neoplasia. However, morphological criteria alone still pose problems for the differential diagnosis and depend upon location and size of liver biopsy as well as doctors' expertise. Therefore, gene expression profiles and more cost-efficient immunohistochemical stainings for glypican 3 (GPC3), heat shock protein 70 (HSP-70), and glutamine synthetase (GS) are recommended by current international clinical practice guidelines to establishing a diagnosis (European Association for the Study of the Liver and European Organisation for Research and Treatment of Cancer 2012). Nevertheless, always a panel of several markers is necessary to achieve a required level of accuracy and a more standardized diagnosis of HCC, but the rate of false negative diagnosis is still high (Di Tommaso et al. 2009; Bruix and Sherman 2011).

Against this background, completely new approaches should be explored that are able to detect various markers in the tissue at once and to use this "fingerprint" for an objective tumor differentiation. Moreover, these innovative approaches could also be used for a classification of the severity of the disease and perhaps even for a risk stratification of the patients. Vibrational spectroscopic techniques such as infrared or Raman spectroscopy are efficient tools for investigating rapidly and non-destructively a wide variety of biochemical compounds in parallel directly in cells or tissue and without any need of labeling. Raman spectroscopy uses the inelastic scattering of visible or near-infrared light to analyze vibrational modes of molecules. One advantage of Raman imaging spectroscopy in comparison with the conventional Raman spectroscopy is high resolution (<1 µm, i.e., subcellular resolution) which allows to determine the spatial distribution of biochemical compounds in tissue sections and to correlate the obtained molecular information with the morphological structures at the very same sample location. Using pattern recognition and classification algorithms, tissue-specific patterns can be discovered and used to develop predictive models. To date, Raman imaging has already been successfully applied in several studies to elucidate the molecular information in colon, cervical, brain, and stomach malignant tissues (Beljebbar et al. 2009; Kamemoto et al. 2010; Bergner et al. 2012; Bergholt et al. 2010). As Raman spectroscopic data will become available for various tumor tissue types, this approach will provide a common disease-wide methodology that can be applied to a variety of clinical questions.

Materials and methods

Tissue collection and sample preparation

The research project was approved by the local medical ethics committee, and written informed consent was obtained from all patients. In the Department of General, Visceral and Vascular Surgery (Jena University Hospital, Germany) surgically resected liver tissue samples were obtained from 23 patients diagnosed with HCC. Immediately after liver resection, tissue samples of the tumor margins were shock frozen in liquid nitrogen and stored at −80 °C. For diagnosis, a histological examination of the resected tissue was performed in the Institute of Pathology (Jena University Hospital, Germany). Table 1 shows the pathological characteristics of these samples.

On the day of Raman measurements, a microtome (Cryostat Leica 3050 S, Leica Biosystems, Germany) was used to prepare two parallel sections of each tissue sample. The first section was mounted onto microscope glass slide and processed with standard hematoxylin and eosin (HE) staining to define the areas of interest for Raman analysis. The second tissue section was deposited onto a calcium fluoride slide (CaF_2; Vacuum-UV quality, Crystal GmbH, Berlin) and transferred to the Raman spectrometer in a vacuum desiccator. According to the pre-defined region of the HE-stained tissue section, Raman images were acquired from the correlating region of the parallel section. After the measurements, the tissue sections on CaF_2 slides remained intact and were also stained with HE, microscopic images were taken and correlated with the obtained Raman maps.

Raman image acquisition

Raman images were collected using a confocal Raman microscope (WITec, Ulm, Germany, Model CRM 2000) with 300 lines/mm grating (blaze wavelength 750 nm) and a 785 nm diode laser as excitation. The laser power was 50 mW. The laser light was focused on the sample with a 50× NA 0.95 objective (EC Epiplan-Apochromat, Zeiss, Germany) coupled to the microscope by a single mode optical fiber. The scattered Raman signal was detected by a back-illuminated deep-depletion CCD camera operating at −65 °C. The Raman system was calibrated to the 520.7 cm^{-1} spectral line of silicon. Raman spectral images of the selected tissue regions were acquired with a measurement area of 75 × 75 $µm^2$ for tumor center and fibrotic regions and a measurement area of 75 × 100 $µm^2$ for the tumor margin. As lateral resolution of the Raman image a 1 µm step size in the spectral region of 3200–200 cm^{-1} was set in the mapping mode of WITec Control. Dark tissues such as kidney and liver have often caused problems for Raman spectroscopy because they are highly absorbing and fluorescent at most visible and NIR wavelengths (Matousek and Stone 2009). Therefore, pre-bleaching of the autofluorescence of 2 s followed by an integration time for each spectrum of 5 s was necessary. The mounting material CaF_2 was chosen to avoid interfering fluorescence background from usually used glass slides.

Table 1 Pathological characteristic of HCC samples

No	Gender	Age	Grade of differentiation	pT	pN	pM	Invasion into lymphatic vessels	Microvascular invasion
1	M	78	Moderate	2	0	0	0	X
2	F	75	Moderate	1	0	0	0	0
3	M	22	Moderate	3	1	X	1	0
4	M	63	Moderate	2	0	X	0	1
5	M	77	Moderate	3	0	X	0	1
6	M	56	X	2	0	X	0	0
7	M	70	Moderate	2	0	0	X	1
8	M	70	Well	X	X	X	X	X
9	M	62	Moderate	2	0	X	0	0
10	F	75	Moderate	1	0	0	0	0
11	M	78	Well	2	0	0	0	X
12	M	65	X	2	X	X	X	1
13	M	71	Moderate	3	0	X	X	1
14	M	86	Moderate	1	0	0	0	0
15	F	51	Moderate	3	0	X	0	0
16	F	81	Poor	3	0	X	0	1
17	M	75	Moderate	3	0	X	0	1
18	M	66	X	3	X	X	X	1
19	F	81	X	1	X	X	0	0
20	M	62	X	X	0	X	X	X
21	M	67	Well	3	0	1	X	1
22	M	62	Poor	4	0	0	X	1
23	F	47	Moderate	3	0	X	0	0

TNM classification for hepatocellular carcinoma: Primary tumor (T): TX—Primary tumor cannot be assessed, T0—No evidence of primary tumor, T1—Solitary tumor without vascular invasion, T2—Solitary tumor with vascular invasion or multiple tumors, none >5 cm, T3a—Multiple tumors >5 cm, T3b—Single tumor or multiple tumors of any size involving a major branch of the portal or hepatic vein, T4—Tumor(s) with direct invasion of adjacent organs other than gallbladder or with visceral peritoneum; Regional lymph nodes (N): NX—Regional lymph nodes cannot be assessed, N0—No regional lymph node metastasis, N1—Regional lymph node metastasis; Distant metastasis (M): M0—No distant metastasis, M1—Distant metastasis

X—unknown, 0—no, 1—yes

Data analysis

The chemometric analysis of the Raman data set of all maps was performed with CytoSpec (CytoSpec Inc., USA), OPUS (Bruker Optik GmbH, Germany) and R (R Core Team 2014).

As an important first step of our data analysis, a data pre-processing was performed. Therefore, interfering signals such as fluorescence, burning artifacts, and present contaminants were determined and removed according to defined upper and lower thresholds for intensity of the CH-stretching region using a quality test included in the CytoSpec software. Artifacts from the recording due to cosmic radiation were removed from the data set by suitable algorithms. Furthermore, baseline correction and vector normalization were applied to the spectral region of 3100–600 cm^{-1}. These spectral pre-processing steps are appropriate for vibrational spectroscopy and were already successfully applied for Raman imaging (Neugebauer et al. 2010; Mazur et al. 2013; Bielecki et al. 2012). A detailed description is published by Lasch (2012).

After that, average spectra from Raman maps of the tumor center and fibrotic tissue region were calculated with the CytoSpec software. In addition, by use of the OPUS software, a difference spectrum of these average spectra was generated by subtraction.

All further data analysis was performed in R using packages hyperSpec (Beleites and Sergo 2014) for import and handling of the data as well as lattice (Sarkar 2008) and ggplot2 (Wickham 2009) for graphical display. For the analyses in R, further preprocessing steps were used: firstly, the uninformative spectral range between 1800 and 2800 cm^{-1} was deleted, and secondly, the spectra were laterally binned 2 × 2 in order to gain signal to noise ratio. Last but not least, the spectra were normalized to the mean intensity of the C–H stretching region between 2800 and 3050 cm^{-1}.

In order to receive a first overview of the inter-individual tissue variability and to detect different chemical components in the tissue samples, all spectral image data were initially subjected to N-FINDR analysis (Winter 1999). N-FINDR is an unsupervised so-called endmember extraction or spectral unmixing algorithm. It is based on the assumptions that the spectra are stemming from a mixture of a known (pre-specified) number of pure component spectra, and that somewhere in the data these pure component spectra ("endmembers") are available. Internally, N-FINDR uses a representation of the data which corresponds to a mixture diagram and the endmembers are determined as corners of that mixture diagram. Since this algorithm does not perform an exhaustive search, the results may be sensitive to the selection of initial endmembers (Plaza and Chang 2005). However, this is typically a symptom that the assumptions are not sufficiently well met, in particular that the specified number of endmembers may have not been appropriate. When replicating the N-FINDR analysis 100 times with randomly chosen initial conditions, we always obtained the same set of endmembers, indicating that the obtained solution is stable. In a second step, for each spectrum the so-called abundances which correspond to the concentrations in the mixture are calculated. Abundances were subjected to a non-negativity constraint as concentrations can take positive values only.

N-FINDR has been used to visualize the variability of Raman spectra within malignant and non-malignant brain tissue regions (Bergner et al. 2012) and within single cells (Hedegaard et al. 2011; Stiebing et al. 2014). Our calculations were carried out using package unmixR (McManus et al. 2013) specifying that 6 endmembers should be extracted. The endmember spectra were compared with Raman spectra of reference substances (proteins, collagen, triglycerides, and cholesterol ester) (De Gelder et al. 2007; Bonifacio et al. 2010; Bonifacio and Sergo 2010; Krafft et al. 2005; Köhler et al. 2009).

As the N-FINDR results indicate a high variability of the biochemical composition within each tissue class (see "Results" section), we chose a nonlinear algorithm for classification. High variability between a comparably small number of patients implies that any kind of spectroscopic classification can become unstable, that is, the model and its predictions change substantially if, e.g., new patients become available in the training data. We meet this difficulty in two ways: Firstly, our choice of classification algorithm is an ensemble model. Ensemble models account for this instability by internally varying the training data base. Briefly, the process is like obtaining the average of multiple noisy measurements—just that here it is the model that is considered "noisy" (Beleites and Salzer 2008). Ensemble models of vibrational spectroscopic data have been successfully used for detection of a variety of diseases (Beleites and Salzer 2008; Menze et al. 2007; Teh et al. 2009; Kallenbach-Thieltges et al. 2013; Ollesch et al. 2013). The random forest algorithm applied here was first described by Breiman (1996). Briefly, decision trees are grown that partition the data according to thresholds for individual wavenumber positions. The "forest" consists of a number of such decision trees which are grown on a randomly chosen ("bootstrapped") subset of spectra and wavenumbers. We used package random Forest (Liaw and Wiener 2002) to grow an ensemble of 500 trees, limiting the leaf node size to not less than 100 spectra. It should be noted that the algorithm is not aware of the hierarchical nature of the data, namely that many spectra are available of each patient. Therefore, the usual out-of-bag classification error estimate (Breiman 1996) cannot be used.

Instead, the performance of the classifier was measured with 101× iterated patient-wise sevenfold cross-validation, (Kohavi 1995; Beleites et al. 2005) which was parallelized using package snow (Tierney et al. 2013). The iterated setup here allows to measure not only the average performance but also the stability of the predictions (Beleites and Salzer 2008).

Results

Comparative analysis of Raman maps

Three regions of interest, including tumor center, fibrotic tissue and tumor margin, of in total 23 tissue samples from 23 patients were measured with Raman imaging spectroscopy, pre-processed, analyzed and correlated with HE-stained microscopy images of the same tissue regions. After the Raman measurements all cryosections were examined a second time by a pathologist and classified into malignant regions of HCC ($n = 12$, 243713 spectra), cancer-free regions of fibrosis ($n = 17$, 330876 spectra) and regions of tumor margin (n = 8, 44709 spectra). Raman spectroscopy allows detecting all substances in the sample at the same time. Despite the strong similarities between the spectra of many biomolecules, there are significant differences in the band positions for different cellular compounds such as lipids and proteins. The chemometric data analysis tools can be used to evaluate precisely the spectral data and find spectral differences for the classification of malignant tumors.

Figure 1 shows the Raman spectra of the main biological components that were found in the tissue sections. The protein spectrum reflects the typical Raman features of the protein backbone and its residues. The CH-stretching vibrations are located between 2800 and 3100 cm^{-1}, the stretching of the C=O of the amide groups (usually referred to as Amide I band) is centered around 1655 cm^{1}, CH_2 deformations are observed at 1450 cm^{-1}. The spectral region

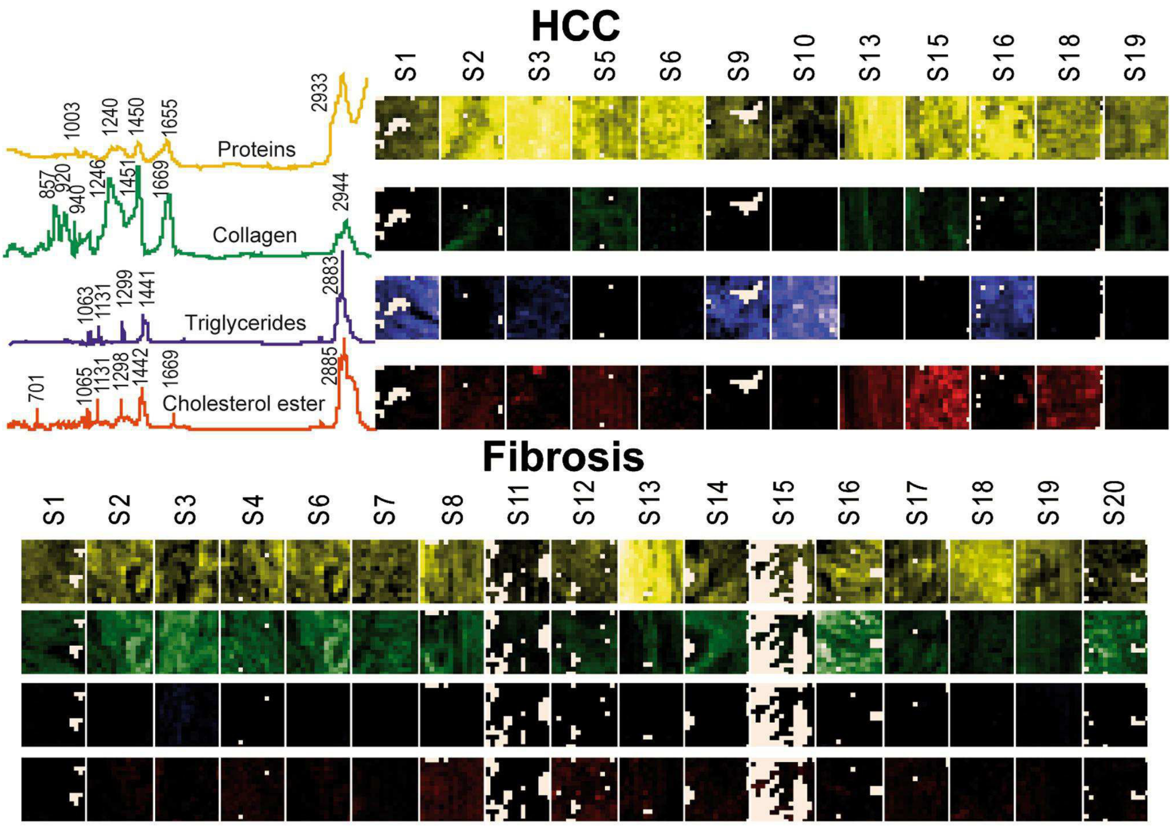


Fig. 1 Abundance plots of HCC regions ($n = 12$) and fibrosis regions ($n = 17$) using the N-FINDR algorithm: proteins (row No 1, *yellow*), collagen (row No 2, *green*), triglycerides (row No 3, *blue*) and cholesterol ester (row No 4, *red*). The Raman spectra of the reference substances were acquired with the same Raman instrument and can be seen in the *upper-left* corner

between 1200 and 1350 cm^{-1} is due to bending vibrations of the amide bonds (usually referred to as Amide II band) and other CH deformations. The small but reproducible band at 1003 cm^{-1} is associated with phenylalanine. The spectrum of collagen is very similar to those other proteins, but exhibits some additional features due to its unique triple helical structure as for instance the bands at 940 and 920 cm^{-1}. The bands of the Raman spectra of triglycerides and cholesterol esters can all be assigned to deformations of the alkyl chains and the steran skeleton.

For each Raman map, the average spectra were calculated and grouped according to the tissue type. Higher band intensities appeared at 2932, 2898, 1654, 1441, 1299, 1261, 1123, 1076, 1001, 953, 923, 851, 748, 721, and 645 cm^{-1} in regions of HCC (Fig. 2a) and 2944, 1661, 1448, 1341, 1257, 1001, 937, 856, 747, and 668 cm^{-1} in regions of fibrosis (Fig. 2b).

The spectral changes of the two tissue types can be analyzed more precisely by difference spectra, which are plotted in Fig. 2c. The positive and negative bands of the difference spectra correspond to the main differences in cancer and fibrotic tissue, respectively. The identified band positions of the intracellular constituents were correlated with reference databases of Raman spectra (De Gelder et al. 2007; Movasaghi et al. 2007). The positive bands relating to regions of HCC showed characteristic bands of lipids or rather saturated and unsaturated fatty acids at the CH-stretching region 2800–3200 cm^{-1} and at the wavelengths of 1655, 1439, 1298, 1087, 1063, and 890 cm^{-1}. The negative bands of the difference spectra relating to regions of fibrosis showed the characteristic features of collagen near 1387, 1343, 1242, 937, 857 cm^{-1}. In addition, in both tissue types (HCC and fibrosis), different protein bands were identified (1677, 1017, 851, 754, 668, 645, 637 cm^{-1}).

Unsupervised unmixing of Raman microspectroscopic images using N-FINDR

Using the N-FINDR algorithm, four endmembers were obtained in all tissue regions that reveal spectral signatures

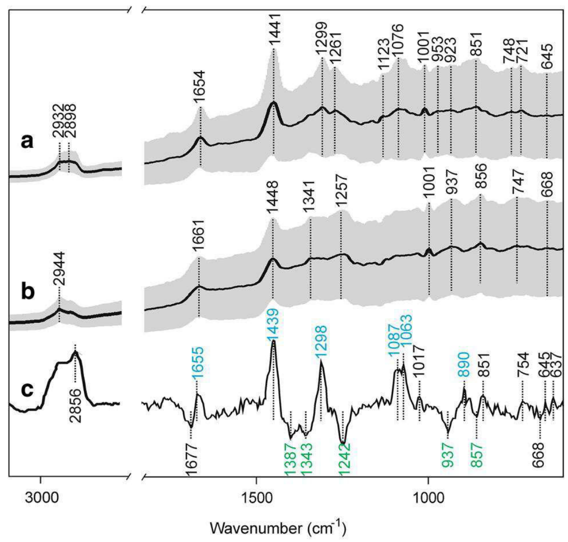


Fig. 2 Average Raman spectra of (**a**) HCC ($n = 12$) and (**b**) fibrosis ($n = 17$) regions in the spectral range of 3100–600 cm^{-1}. **c** Difference spectrum (×5) of HCC versus fibrosis. Identified Raman bands are indicated by the corresponding wavenumber (in cm^{-1}). The numbers *colored in blue* can be assigned to lipids and *colored in green* to collagen. *Shaded areas* of the spectra a and b represents the standard deviations

of proteins, triglycerides, collagen, and cholesterol esters. The color-coded Raman maps of the obtained endmembers in regions of HCC and fibrotic tissue can be seen in Fig. 1. HCC tissue mostly included proteins or triglycerides as main components and only small amount of collagen or cholesterol esters were found. The fibrotic tissue consists of collagen, proteins, and small amount of cholesterol ester. No triglyceride signatures were found in the Raman maps of fibrotic regions. These results correlate well with the typical chemical composition of HCC and fibrotic tissue (George and Chandrakasan 2000). Although none of these identified endmembers showed a significant difference in the visualized Raman maps, the results of the analysis demonstrates the heterogeneity and complexity of our data set. Because the N-FINDR analysis resulted in a high inter-individual variability of the tissue samples and a heterogeneous distribution of typical chemical components of HCC and fibrotic tissue was identified, the data analysis was supplemented by a nonlinear classification algorithm.

Random forest classification model and variables importance plot

In order to identify relevant signatures and to classify malignant and non-malignant liver tissue, a random forest classification model was applied to the complete data set of regions of HCC and fibrosis. A 101× iterated sevenfold cross-validation estimating classifier performance characteristics resulted in an accuracy of 86 % (sensitivity of 76 % and specificity of 93 %) to predict HCC. The spectral bands used by the random forest classifier were ranked by importance. In Fig. 3, the variables importance plot of the random forests is demonstrated. Based on the importance intensities of variables, the ten most discriminating bands were determined at 2895, 2856, 1439, 1298, 1080, 1063, 1023, 937, 920, and 719 cm^{-1}.

Prediction maps

In addition, the calculated cross-validated classification model was used to generate color-coded prediction maps

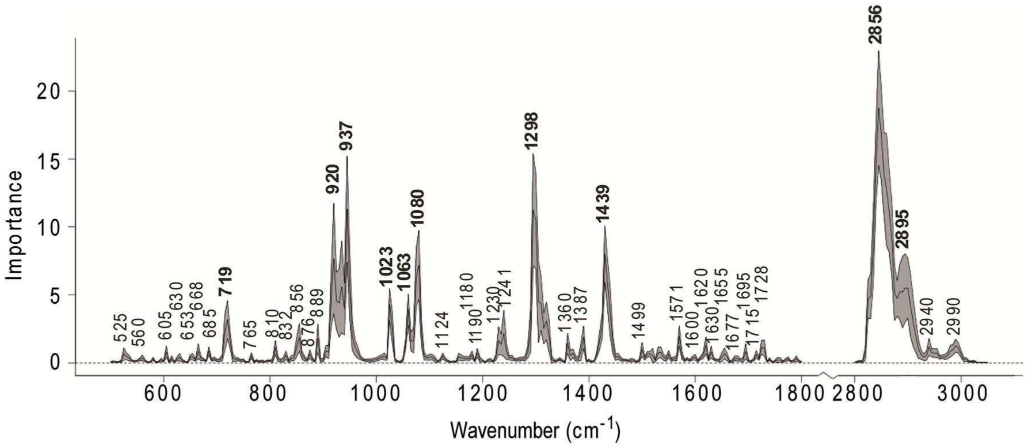


Fig. 3 Variables importance plot. The importance of variables is ranked in this plot according to the height of the bands. *Shaded areas* represent the standard deviations

of unknown regions of tumor margins. Consequently, each Raman map of the region in-between cancer and fibrosis was predicted and assembled as a color-coded image (Fig. 4). The color-intensity scale illustrates the positive decisions of the tree predictors in percentage. The dark blue color is consistent with 0 % positive decisions for HCC and the color red is consistent with 100 % positive decisions for HCC. In total, 8 unknown regions of tumor margin were predicted into HCC and fibrosis. By correlating the color intensity of the prediction maps with the corresponding tissue architecture of the same sample location in the HE-stained tissue section, the classifier discriminates between regions of HCC and fibrosis with high accuracy using a cut-off value of 50 %.

Discussion

To develop innovative and effective diagnostic methods and therapeutic strategies, it is crucial to first identify the target molecules of the underlying disease. By the combination of highly specific Raman spectroscopic information and the spatial resolution of microscopic imaging, the novel Raman imaging technology is a very promising and powerful tool, especially in terms of sensitivity and specificity, to detect and localize a variety of unknown molecules at the cellular level in a label-free tissue manner. Thus, this technique provides spectral information at molecular level of the underlying tissue structure.

Following on a previous study, Tolstik et al. (2013) in which we investigated liver cancer cell lines of HCC (HepG2) and adenocarcinoma (SK-Hep1) by Raman imaging, we analyzed in this study a number of different regions of human HCC and fibrosis in sections of cirrhotic liver tissue with this spectroscopic technique. For this approach, acquired Raman imaging data of predefined malignant and non-malignant tissue regions were subjected to a comparative analysis and were used to establish a classification model based on a random forest algorithm. A cross-validation technique was performed to estimate the prediction accuracy of this model. Furthermore, we applied the developed algorithm to Raman imaging data of tumor margins and the adjacent fibrotic tissue to verify the predictive performance of the resulted classifier.

For diagnostic use, the obtained Raman pattern was applied in the above-described random forests model. Estimating classifier performance characteristics, a cross-validation was performed that resulted in a sensitivity of 76 % [95 % c.i. using an effective sample size of 12 patients (Dorai-Raj 2014): 49–93 %] and specificity of 93 % (73–99 %). The prediction maps of the tumor margin regions (Fig. 4) visualize the prediction accuracy of the classification algorithm in form of a color-coded map. Besides the tumor area, even minimal amounts of tumor cells dispersed over the tumor margin might be detectable in the tissue section using this high-resolution detection method. For some other tumor entities, the assessment of tumor margin by Raman spectroscopy has already been successfully demonstrated for in vivo (using an optical fiber Raman probe) and ex vivo experimental approaches (Hughes et al. 2010; Haka et al. 2006). The results of these studies and our findings indicate that vibrational spectroscopy techniques, such as Raman spectroscopy, have an enormous potential for the future to assess intraoperatively the tumor margin of HCC in real time during liver resection and with potential application in other malignancies.

In addition to these results, the random forest model provides the so-called variable importance measure, which is shown in Fig. 3. Here, the measure depends on the increase in classification errors if the wavenumber channel in question is not available (Liaw and Wiener 2002). It is important to realize that the vibrational spectrum of one substance consists of a number of different bands due to the different vibrational modes that are possible within the molecule. In principle, all these bands are equivalent to each other. Thus, while high importance implies that the band in question and therefore also the biochemical substance giving raise to this band is important for the distinction of HCC from fibrosis, we cannot conclude anything from low importance values: Not allowing the use of one specific band (say, a symmetric stretching band), the random forest may just switch over to use one of the other vibrations (e.g., the antisymmetric stretching or a deformation) of the molecule in question.

The ten most important spectral channels were identified at the wavenumbers 2856, 1298, 937, 920, 1439, 1080, 2895, 1023, 1063, and 719 cm^{-1} (ranked in order of importance). By comparison of these most important bands with the difference spectrum of malignant and non-malignant tissue regions, four bands at the wavelength 2856, 1298, 1439, and 1063 cm^{-1} were assigned to regions of HCC, whereas only one band (937 cm^{-1}) among the ten most important bands corresponds to regions of fibrosis (Table 2, column 1, 2). In addition to the ten most important variables, two other Raman bands 1655 and 889 cm^{-1} of the variables importance plot were positive (HCC) and five other important variables at 1677, 1387, 1242, 856, and 668 cm^{-1} were negative (fibrosis) in the difference spectrum. The Raman bands at 2895, 1080, 1023, 920, and 719 cm^{-1} of the importance plot could not be attributed to a specific tissue region. However, the bands at 1080 and 719 cm^{-1} were assigned to the phospholipid phosphatidylcholine. The occurrence of the choline band at 719 cm^{-1} was not expected because a difference band was not evident (Fig. 2). The role of phosphatidylcholine in detection of fibrosis and HCC will be studied in more detail, after more specimens are available.

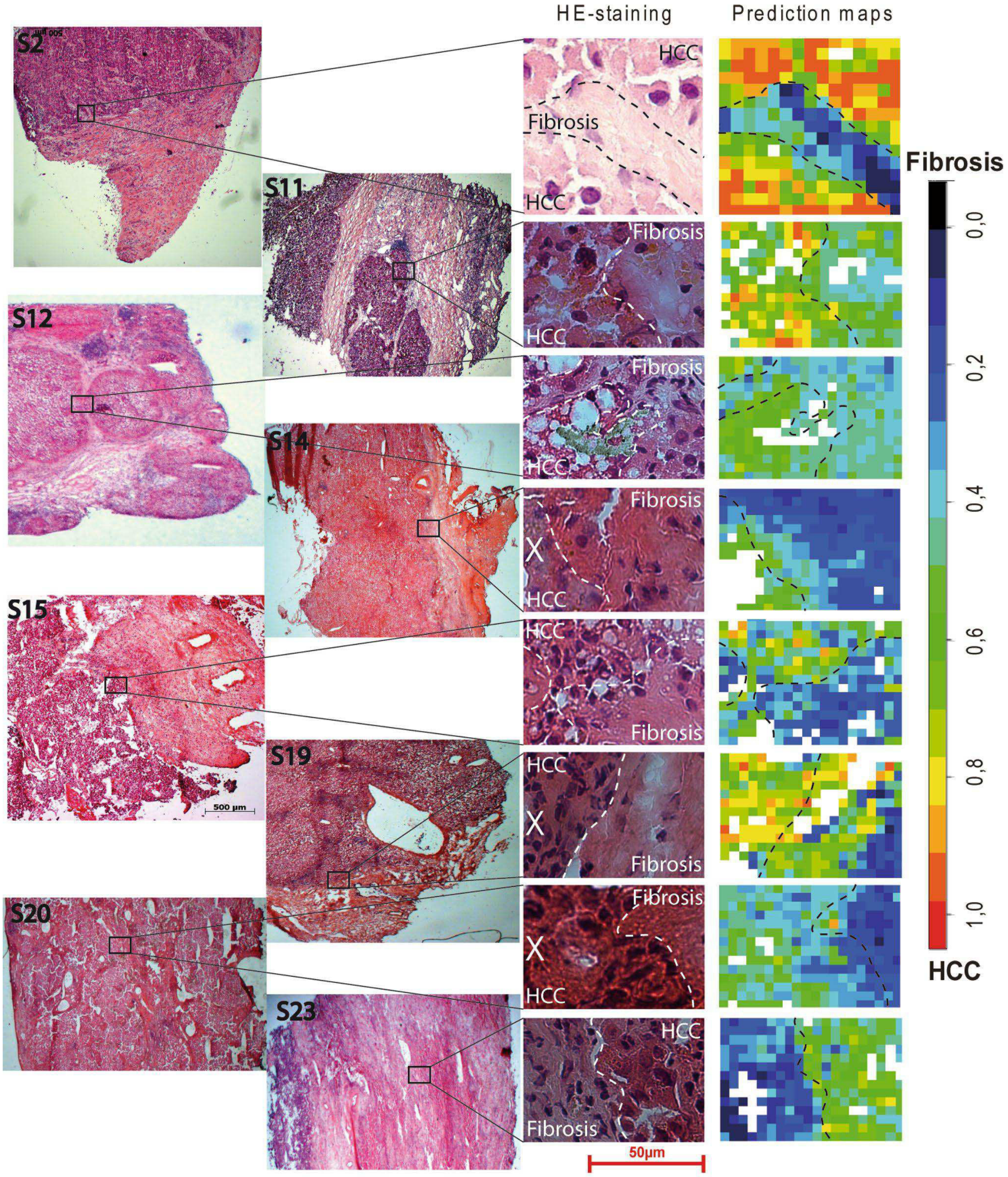


Fig. 4 HE-stained tissue section and *color*-coded prediction *maps* of tumor margin regions. The *color*-intensity scale illustrates the *positive* decisions of the random forest predictors. X—parallel tissue sections

Using independent reference databases of Raman spectra for biological molecules (De Gelder et al. 2007b; Bonifacio et al. 2010; Bonifacio and Sergo 2010; Krafft et al. 2005; Köhler et al. 2009) (Table 2, column 3, 4), the most discriminating bands detected in HCC tissue regions could be correlated with Raman signature of the palmitic

Table 2 Correlation of important Raman bands with independent reference databases and Raman data of cell lines

Difference spectrum of tissue regions (see Fig. 2c)		VIP of the RFM (see Fig. 3)		Reference database 1 (De Gelder et al. 2007)		Reference databases 2 (Bonifacio et al. 2010; Bonifacio and Sergo 2010; Krafft et al. 2005; Köhler et al. 2009)		Difference spectrum of cell lines (HepG2 vs. SK-Hep1)
Positive bands (HCC)	Negative bands (Fibrosis)	Rank	Raman bands (in cm^{-1})	Raman bands (in cm^{-1})	Product	Bands (in cm^{-1})	Assignment	Discriminating bands (in cm^{-1})
–	–	7	2895	–		–		2898
2856	–	1	2856	–		–		2854
–	1677		1677	–		–		–
1655	–		1655	1655	Amide I band (proteins), ν(C=C) stretching vibrations (unsaturated fatty acids)	1659	Lipids (Krafft et al. 2005; Köhler et al. 2009)	1655
1439	–	5	1439	1438 (vs)	Palmitic acid	1440	Lipids (Krafft et al. 2005; Köhler et al. 2009)	1440
–	1387		1387	1389 (m)	L-Proline	–		–
–	1343		–	1343 (s, sh)	L-Valine	–		–
1298	–	2	1298	1296 (vs)	Palmitic acid	1299	Lipids (Krafft et al. 2005; Köhler et al. 2009)	1304
–	1242		1242	1240 (m)	L-Proline	1246	Collagen (Bonifacio et al. 2010; Bonifacio and Sergo 2010)	–
1087	–		–	1087 (s)	Phosphate	1087	Lipids (Krafft et al. 2005; Köhler et al. 2009)	–
–	–	6	1080	1080 (m)	Oleic acid	–		1078
1063	–	9	1063	1063 (vs)	Palmitic acid, myristic acid	1064	Lipids (Krafft et al. 2005; Köhler et al. 2009)	1059
–	–	8	1023	1023 (m, sh)	Oleic acid	–		–
1017	–		–	–		–		–
–	937	3	937	930 (m)	L-Proline	940	Collagen (Bonifacio et al. 2010; Bonifacio and Sergo 2010)	–
–	–	4	920	916 (vs)	L-Proline	920	Collagen (Bonifacio et al. 2010; Bonifacio and Sergo 2010)	–

Table 2 continued

Difference spectrum of tissue regions (see Fig. 2c)		VIP of the RFM (see Fig. 3)		Reference database 1 (De Gelder et al. 2007)		Reference databases 2 (Bonifacio et al. 2010; Bonifacio and Sergo 2010; Krafft et al. 2005; Köhler et al. 2009)		Difference spectrum of cell lines (HepG2 vs. SK-Hep1)
Positive bands (HCC)	Negative bands (Fibrosis)	Rank	Raman bands (in cm^{-1})	Raman bands (in cm^{-1})	Product	Bands (in cm^{-1})	Assignment	Discriminating bands (in cm^{-1})
890	–		889	889 (m)	Vaccenic acid	892	Triglycerides (Krafft et al. 2005)	–
				890 (w, m)	Stearic acid (w)			
				893 (m)	Palmitic acid, myristic acid			
–	857		856	857 (vs)	L-Proline	857	Collagen (Bonifacio et al. 2010; Bonifacio and Sergo 2010)	–
851	–		–	850 (s, sh)	L-Proline	–		–
754	–		–	–		–		–
–	–	10	719	719	Choline	720	Lipids (Krafft et al. 2005; Köhler et al. 2009)	717
–	668		668	–		–		–
645	–		–	–		–		–
637	–		–	–		–		–

acid at 1438, 1296, 1063, and 893 cm^{-1}. In contrast to this, the most discriminating band at 937 cm^{-1} detected in fibrotic tissue regions was assigned to collagen. Furthermore, additional bands of the variables importance plot that simultaneously appeared in the difference spectrum were attributed to unsaturated or saturated fatty acids (vaccenic acid near 889 cm^{-1} and oleic acid near 1080, 1023 cm^{-1}, or stearic acid near 890 cm^{-1}) in regions of HCC and other Raman bands of collagen (1246, 940, 820, and 857 cm^{-1}) in regions of fibrosis (Table 2) (De Gelder et al. 2007).

The results of this study are in good agreement with our previously obtained spectral data and classification algorithm of liver cancer cell lines using the Raman imaging technique Tolstik et al. (2013) (Table 2, column 5) and are in accordance with the current state of research relating to the "de novo lipogenesis" in cancer (Menendez and Lupu 2007; Currie et al. 2013). In a very recently published study, Tolstik et al. (2013) we were able to discriminate and classify liver cancer cells, including cell lines of HCC (HepG2) and adenocarcinoma of the liver (SK-Hep1), and their proliferative stages with Raman imaging. The main detected differences in Raman spectra were attributed to a higher expression of fatty acids in the HCC cell line and in the exponential phase of tumor cell growth.

Currently, it is already known that aberrant lipid biosynthesis is an important mechanism in pathogenesis of liver diseases and HCC. A small-scale study of 10 HCC tissue samples published in 2005 described already an elevated expression of mRNA for fatty acid synthase (FASN), acetyl-CoA carboxylase (ACAC) and ATP citrate lyase (ACLY) in HCC tissue (Yahagi et al. 2005; Calvisi et al. 2011; Gao et al. 2006). In 2011, a large-scale study demonstrated a progressive upregulation from non-malignant liver tissue toward the HCC of all relevant lipogenic enzymes, which are involved in the fatty acid synthesis (FASN, ACAC, ACLY, ME, and SCD1), the cholesterol biosynthesis (SREBP2, HMGCR, MVK, and SQS) and their upstream inductors (chREBP, SREBP1, and LXR-β). This latter study showed as well that a higher expression of these enzymes correlate with an activation of a lipogenic pathway (AKT-mTORC1-RPS6) and the clinical aggressiveness, characterized by shorter and longer patient survival (Calvisi et al. 2011).

These results have recently been substantiated by another published study that identified lipid biomarkers associated with progression of HCC and patient outcomes using a combination of gene expression and metabolic profile analysis. In this study, the authors identified 28 metabolites and 169 genes associated with aggressive HCC. Lipid metabolites of stearoyl-CoA-desaturase (SCD) activity were associated with aberrant palmitate signaling in aggressive HCC samples. Expression of gene products associated with these metabolites, including SCD, were associated independently with survival times and tumor recurrence in the test and validation sets. Combined expression of SCD and α-fetoprotein were associated with outcomes of patients with early-stage HCC. Levels of palmitic acid, the product of SCD activity, were increased in aggressive HCCs; palmitic acid increased migration and invasion of cultured HCC cells and colony formation by HCC cells. HCC cells that expressed small interfering RNA against SCD had decreased cell migration and colony formation in culture and reduced tumorigenicity in mice (Budhu et al. 2013).

Despite of the promising data demonstrated by this study, some limitations should be mentioned. First, the size of the sample collective of the study is limited. This is attributed to the fact that in this study frozen tissue samples predefined for HCC were used to avoid potential artifacts caused by sample modification by some kind of fixation. Clearly, before considering the use of a Raman imaging algorithm for molecular classification of HCC patients suited for confirmation of diagnosis or risk stratification, future large-scale studies will be required to confirm our findings and to correlate the results of the spectral data with testing results of approved methods. Second, Raman measurements were taken from ex vivo tissue.

Conclusion

In summary, we have demonstrated that the use of Raman imaging can provide a classification and prediction model for HCC in cirrhotic liver tissue. Furthermore, the most discriminating bands detected in HCC tissue were identified as palmitic acid. Increased levels of saturated palmitic acid, as well as its activating enzyme (SCD), were associated with aggressive HCC, characterized by survival times, tumor recurrence and outcomes of patients with early-stage HCC plus increased migration and invasion of cultured HCC cells (Budhu et al. 2013). Due to the fact that aberrant lipid biosynthesis, upregulation of relevant lipogenic enzymes and activation of lipogenic pathways with a resulting imbalance of lipogenic components are important factors in pathogenesis and aggressiveness of HCC, effective diagnostic methods, including Raman spectroscopy, could be helpful to identify these lipid biomarkers. Even a future use of Raman spectroscopy for intraoperative risk stratification would be conceivable that may cause a change in the surgical procedure. Although subject to validation of our findings in larger HCC cohorts, the demonstrated results affirm that a classifier based on Raman imaging data could provide additional molecular information for diagnostics and risk stratification in HCC.

Acknowledgments We are grateful to all co-workers from the Gastroenterology Laboratory of the Jena University Hospital and the Spectroscopy/Imaging group in IPHT for their support and help. Financial support from the "Deutsche Forschungsgemeinschaft (DFG)", the "Europäischer Fonds für Regionale Entwicklung (EFRE)" and the "Thüringer Ministerium für Bildung, Wissenschaft und Kultur (TMBWK)" are gratefully acknowledged. CB Acknowledges Funding by the BMBF via project RamanCTC (AZ 13N12685).

Conflict of interest We declare that we have no conflict of interest.

References

Beleites C, Salzer R (2008) Anal Bioanal Chem 390:1261–1271

Beleites C, Sergo V (2014) HyperSpec: a package to handle hyperspectral data sets in R, version 0.98-20140612. http://hyperspec.r-forge.r-project.org

Beleites C, Baumgartner R, Bowman C, Somorjai R, Steiner G, Salzer R, Sowa MG (2005) Chemom Intell Lab Syst 79:91–100

Beljebbar A, Bouché O, Diébold MD, Guillou PJ, Palot JP, Eudes D, Manfait M (2009) Crit Rev Oncol Hematol 72:255–264

Bergholt MS, Zheng W, Lin K, Ho KY, Teh M, Yeoh KG, So JBY, Huang Z (2010) Analyst 135:3162–3168

Bergner N, Krafft C, Geiger KD, Kirsch M, Schackert G, Popp J (2012) Anal Bioanal Chem 403:719–725

Bielecki C, Bocklitz TW, Schmitt M, Krafft C, Marquardt C, Gharbi A, Knösel T, Stallmach A, Popp J (2012) J Biomed Opt 17:076030

Bonifacio A, Sergo V (2010) Vib Spectrosc 53:314–317

Bonifacio A, Beleites C, Vittur F, Marsich E, Semeraro S, Paoletti S, Sergo V (2010) Analyst 135:3193–3204

Breiman L (1996) Mach Learn 24:123–140

Bruix J, Sherman M (2011) Hepatology 53:1020–1022

Budhu A, Roessler S, Zhao X, Yu Z, Forgues M, Ji J, Karoly E, Qin L-X, Ye Q-H, Jia H-L, Fan J, Sun H-C, Tang Z-Y, Wang XW (2013) Gastroenterology 144:1066–1075 e1

Calvisi DF, Wang C, Ho C, Ladu S, Lee SA, Mattu S, Destefanis G, Delogu S, Zimmermann A, Ericsson J, Brozzetti S, Staniscia T, Chen X, Dombrowski F, Evert M (2011) Gastroenterology 140:1071–1083

Currie E, Schulze A, Zechner R, Walther TC, Farese RV Jr (2013) Cell Metab 18:153–161

De Gelder J, De Gussem K, Vandenabeele P, Moens L (2007a) J Raman Spectrosc 38:1133–1147

De Gelder J, De Gussem K, Vandenabeele P, Moens L (2007b) J Raman Spectrosc 38:1133–1147

Di Tommaso L, Destro A, Seok JY, Balladore E, Terracciano L, Sangiovanni A, Iavarone M, Colombo M, Jang JJ, Yu E, Jin SY, Morenghi E, Park YN, Roncalli M (2009) J Hepatol 50:746–754

Dorai-Raj S (2014) Binom: binomial confidence intervals for several parameterizations, version 1.1-1. http://CRAN.R-project.org/package=binom

European Association for the Study of the Liver and European Organisation for Research and Treatment of Cancer (2012) J Hepatol 56:908–943

Gao Y, Lin L-P, Zhu C-H, Chen Y, Hou Y-T, Ding J (2006) Cancer Biol Ther 5:978–985

George J, Chandrakasan G (2000) Clin Biochem 33:563–570

Haka AS, Volynskaya Z, Gardecki JA, Nazemi J, Lyons J, Hicks D, Fitzmaurice M, Dasari RR, Crowe JP, Feld MS (2006) Cancer Res 66:3317–3322

Hedegaard M, Matthäus C, Hassing S, Krafft C, Diem M, Popp J (2011) Theor Chem Acc 130:1249–1260

Hughes OR, Stone N, Kraft M, Arens C, Birchall MA (2010) Head Neck 32:1544–1553

Iavarone M, Colombo M (2013) Clin Liver Dis 17:375–397

Kallenbach-Thieltges A, Großerüschkamp F, Mosig A, Diem M, Tannapfel A, Gerwert K (2013) J Biophotonics 6:88–100

Kamemoto LE, Misra AK, Sharma SK, Goodman MT, Luk H, Dykes AC, Acosta T (2010) Appl Spectrosc 64:255–261

Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the 14th international joint conference on artificial intelligence. Morgan Kaufmann, USA, pp 1137–1145

Köhler M, Machill S, Salzer R, Krafft C (2009) Anal Bioanal Chem 393:1513–1520

Krafft C, Neudert L, Simat T, Salzer R (2005) Spectrochim Acta A Mol Biomol Spectrosc 61:1529–1535

Lasch P (2012) Chemometr Intell Lab Syst 117:100–114

Liaw A, Wiener M (2002) R News 2:18–22

Matousek P, Stone N (2009) Analyst 134:1058–1066

Mazur AI, Monahan JL, Miljković M, Laver N, Diem M, Bird B (2013) J Biophotonics 6:101–109

McManus C, Fuller S, Hanson B, Beleites C (2013) UnmixR: spectral unmixing methods, version 0.1-6. http://github.com/Chathurga/unmixR

Menendez JA, Lupu R (2007) Nat Rev Cancer 7:763–777

Menze BH, Petrich W, Hamprecht FA (2007) Anal Bioanal Chem 387:1801–1807

Movasaghi Z, Rehman S, Rehman IU (2007) Appl Spectrosc Rev 42:493–541

Neugebauer U, Clement JH, Bocklitz T, Krafft C, Popp J (2010) J Biophotonics 3:579–587

Ollesch J, Drees SL, Heise HM, Behrens T, Brüning T, Gerwert K (2013) Analyst 138:4092–4102

Plaza A, Chang C-I (2005) An improved N-FINDR algorithm in Implementation, algorithms and technologies for multispectral, hyperspectral, and ultraspectral imagery XI. Proc SPIE 5806:298–306. doi:10.1117/12.602373

R Core Team (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Sarkar D (2008) Lattice: multivariate data visualization with R. Springer, New York

Stiebing C, Matthäus C, Krafft C, Keller A-A, Weber K, Lorkowski S, Popp J (2014) Complexity of fatty acid distribution inside human macrophages on single cell level using Raman micro-spectroscopy. Anal Bioanal Chem. doi:10.1007/s00216-014-7927-0

Teh SK, Zheng W, Lau DP, Huang Z (2009) Analyst 134:1232–1239

Tierney L, Rossini AJ, Li N, Sevcikova H (2013) Snow: simple network of workstations, version 0.3-13. http://CRAN.R-project.org/package=snow

Wickham H (2009) ggplot2: elegant graphics for data analysis. Springer, New York. ISBN 978-0387981406. http://had.co.nz/ggplot2/book

Winter ME (1999) N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data. Proc SPIE 3753:266–275

Tolstik T, Marquardt C, Matthäus C, Bergner N, Bielecki C, Krafft C, Stallmach A, Popp J (2013) Discrimination and classification of liver cancer cells and proliferation states by raman spectroscopic imaging. Anal (submitted)

Yahagi N, Shimano H, Hasegawa K, Ohashi K, Matsuzaka T, Najima Y, Sekiya M, Tomita S, Okazaki H, Tamura Y, Iizuka Y, Ohashi K, Nagai R, Ishibashi S, Kadowaki T, Makuuchi M, Ohnishi S, Osuga J, Yamada N (2005) Eur J Cancer 41:1316–1322

Publication 3:

MALDI imaging-based classification of hepatocellular carcinoma and non-malignant lesions in fibrotic liver tissue

---

Claudio Marquardt, Tatiana Tolstik, Christiane Bielecki, Roland Kaufmann,
Anna C Crecelius, Ulrich S Schubert, Utz Settmacher, Andreas Stallmach, Olaf Dirsch

Published in:

Zeitschrift für Gastroenterologie, **2015 Jan**; 53(1):33-39.

IF (2014): 1.052.

As the second method for molecular imaging of the biological specimens MALDI IMS for proteomic analysis of liver cancer was applied. Therefore, the cryosections with malignant HCC regions (n = 25) and non-malignant fibrotic (n = 11) and regenerative nodules (n=8) regions in hepatic specimens were measured. Overall 116 protein peaks were detected in measured liver samples. After staining of the investigated tissue sections with H&E we were able to correlate mass spectral information with histology. Based on that, an overexpression of previously unknown proteomic patterns at 6222, 6274, 6647 and 6853 m/z was detected in HCC regions. Furthermore, based on most significant three protein peaks (6274, 6647 and 6222 m/z) a classification model was developed and resulted in 90 % of sensitivity, specificity and accuracy for prediction of previously unknown HCC regions in cirrhotic liver tissue. This work showed high potential of MALDI IMS for identification of novel protein biomarkers for HCC diagnostics and algorithm for its classification in contrast to unknown liver lesions.

MALDI imaging-based classification of hepatocellular carcinoma and non-malignant lesions in fibrotic liver tissue

Klassifikation des hepatozellulären Karzinoms und nicht-maligner Läsionen im fibrotischen Lebergewebe mittels bildgebender Massenspektrometrie

Authors C. Marquardt[1], T. Tolstik[1], C. Bielecki[1], R. Kaufmann[2], A. C. Crecelius[3, 4], U. S. Schubert[3, 4], U. Settmacher[2, 5], A. Stallmach[1], O. Dirsch[6]

Affiliations Affiliation addresses are listed at the end of the article.

Schlüsselwörter
- hepatozelluläres Karzinom
- bildgebende Massenspektrometrie
- MALDI IMS
- Leberkrebs
- Proteomik

Key words
- hepatocellular carcinoma
- imaging mass spectrometry
- MALDI IMS
- liver cancer
- proteomics

received 9.3.2014
accepted 25.9.2014

Bibliography
DOI http://dx.doi.org/10.1055/s-0034-1385398
Z Gastroenterol 2015; 53: 33–39 © Georg Thieme Verlag KG Stuttgart · New York · ISSN 0044-2771

Correspondence
Claudio Marquardt
Klinik für Innere Medizin IV, Universitätsklinikum Jena
Erlanger Allee 101
07747 Jena
Germany
Tel.: ++ 49/36 41/9 32 46 39
Fax: ++ 49/36 41/9 32 42 22
claudio.marquardt@med.uni-jena.de

Zusammenfassung

Die histopathologische Differenzierung nodulärer Läsionen in der zirrhotischen Leber ist für Pathologen schwierig, insbesondere in Biopsiematerial. Es werden daher neue Gewebemarker benötigt, um die histopathologische Entscheidung zu stützen. Die bildgebende Massenspektrometrie (Matrix-assisted laser desorption/ionization imaging mass spectrometry – MALDI-IMS) ist eine leistungsfähige Technologie, die es ermöglicht, Proteine direkt im Gewebe zu detektieren und zu lokalisieren, und es ggf. erlaubt, die Diagnose eines hepatozellulären Karzinoms (HCC) zu bestätigen. In dieser Arbeit wurden proteomische Muster in Kryostatschnitten von Läsionen des HCC (n = 15) und von nicht-malignem fibrotischem Lebergewebe (n = 11) mittels MALDI-IMS untersucht. Durch die Korrelation der erzeugten massenspektrometrischen Bilder mit der Histologie des entsprechenden Gewebeschnitts konnten 4 Proteine mit den m/z-Werten 6274, 6647, 6222 und 6853 identifiziert werden, die eine signifikant höhere Expression im HCC-Gewebe im Vergleich zur Fibrose aufweisen. Basierend auf 3 dieser Proteine mit der höchsten Signifikanz wurde ein Klassifikationsmodell erstellt und an einer unabhängigen Gruppe von Gewebeproben des HCC (n = 10) und benigner Regeneratknoten (n = 8) validiert. Mit diesem Klassifikationsmodell konnte eine verlässliche Vorhersage zwischen benignen und malignen Läsionen im fibrotischen Lebergewebe mit einer Sensitivität und Spezifität von 90 % getroffen werden. Diese identifizierte MALDI-IMS-Proteomsignatur kann zukünftig hilfreich sein, um eine endgültige und objektive Erstellung der Diagnose zu vereinfachen und so das Risiko von Verzögerungen bei der Diagnosebestimmung sowie der Therapieeinleitung von HCC-Patienten zu minimieren.

Abstract

Histopathologic differentiation of nodular lesions in cirrhotic liver is difficult even for experienced hepatopathologists especially regarding diagnosis of hepatocellular carcinoma (HCC) in biopsies. For this reason, new tissue markers are needed to reinforce histopathologic decision-making. With advances in molecular techniques, proteomic analysis may help to confirm the diagnosis of HCC. Matrix-assisted laser desorption/ionization imaging mass spectrometry (MALDI IMS) is a powerful technology which allows to determine and to localize proteins directly in tissue sections. Using MALDI IMS proteomic patterns of cryosections with lesions of HCC (n = 15) and non-malignant fibrotic liver tissue (n = 11) were investigated to establish a classification model of HCC, which was validated in an independent set of tissue to distinguish HCC (n = 10) from regenerative nodules (n = 8). By correlating generated mass spectrometric images with the histology of the tissue sections we found that the expression of 4 proteins as indicated by m/z 6274, m/z 6647, m/z 6222 and m/z 6853 was significantly higher in HCC tissue than in non-tumorous liver tissue. The generated classification model based on the most significant 3 differentially expressed proteins allowed a reliable prediction of benign and malignant lesions in fibrotic liver tissue with a sensitivity and specificity of 90 % in the validation set. The identified MALDI IMS proteomic signature can be diagnostically helpful to allow simplifying the diagnostic process and minimize the risks of delays in establishing the objective final diagnosis and initiating treatment of patients with HCC.

Introduction

Hepatocellular carcinoma (HCC) is the sixth most common type of cancer worldwide, and its incidence is increasing [1]. In over 80 % of the cases, HCC develops in chronic liver disease [2]. Major causes of HCC in Europe and the USA are hepatitis C and alcoholic liver disease, whereas in Asia hepatitis B represents the main risk factor for developing HCC [1, 3]. Liver cirrhosis is regarded as a precancerous lesion independent of the etiology of this end stage liver disease [4]. Depending on the etiology the annual risk of developing HCC in cirrhosis varies between 1 and 6 % [2]. Despite advances in treatment, potentially curative resection, liver transplantation or percutaneous treatment can be applied to only 30 % of all patients with HCC [5].

Diagnosis of HCC using cytohistological criteria according to the International Consensus Group for Hepatocellular Neoplasias is often difficult even for experienced hepatopathologists [6]. Besides morphological characteristics a panel of several immunohistochemical markers, such as glypican 3 (GPC3), heat shock protein 70 (HSP-70), and glutamine synthetase (GS) might be helpful, but the rate of false negative diagnosis is still high [6, 7]. New molecular markers could improve the accuracy of the histopathological diagnosis.

So far, the molecular mechanisms of hepatocarcinogenesis and the molecular changes in HCC are not fully understood. Additional molecular information obtained by proteomic analysis may facilitate early diagnosis and identify new targets for specific therapeutic approaches. For several years now, matrix assisted laser desorption/ionization imaging mass spectrometry (MALDI IMS) has found widespread use in the analysis of biomedical samples and represents a promising technology for visualizing the spatial distribution of a variety of different unknown analytes in label-free tissue sections including peptides, proteins, lipids and drug metabolites. By the use of recognition and classification algorithms tissue-specific patterns can be discovered in order to develop an independent prediction model.

MALDI IMS was already successfully applied to detect and identify different biomarkers for a better molecular understanding, classification and prognosis estimation or therapy response of various diseases primarily focusing on malignant diseases but also in the field of gastroenterology and hepatology [8 – 31].

The objective of this study was to identify new specific tissue-based molecular patterns of HCC by means of MALDI IMS in order to discover potential protein biomarkers for identifictaion and targeted treatment of HCC.

Materials and Methods

Ethics statement

The study was approved by the local Medical Ethics Committee (Ethics Committee of the Friedrich Schiller University, Jena, Germany). Written informed consent was obtained from all patients.

Sample collection and tissue specimens

Primary HCC were either resected or explanted between 2003 and 2009 in the Department for General, Visceral and Vascular Surgery (University Hospital of the Friedrich Schiller University, Jena, Germany). After resection or explantation tumor tissue samples were embedded in Tissue-Tek O. C. T. (Optimal Cutting Temperature) compound, immediately snap-frozen, and subsequently stored at −80 °C. In addition, paraffin sections were prepared. Histological diagnosis of the tumors was performed by the Institute of Pathology (**Table 1**). Cryosections were assessed by an experienced hepatopathologist by examining the tissue architecture using a light microscope in a blinded manner and classified histologically into regions of HCC (n = 25), fibrosis (n = 11) and regenerative nodules (n = 8).

Tissue sample preparation

Frozen tissue samples of HCC were cut in a cryostat at −20 °C into 10 µm thick slices and directly deposited onto electrically conductive indium-tin oxide (ITO)-coated and optically transparent glass slides (Bruker Daltonik GmbH, Bremen, Germany). Tissue sections were incubated twice for 1 min in 70 % ethanol (HPLC grade) and 96 % ethanol (HPLC grade) and in between one washing step in distilled water was performed. After drying the section in a vacuum desiccator the slides were marked with dots of liquid white-out depicting points of reference for the later following overlay of mass spectrometric and microscope images (co-registration). Subsequently, all tissue sections received a standardized matrix deposition protocol by spraying very fine droplets of an UV-absorbing matrix solution using an automated spraying device (ImagePrep station, Bruker Daltonik GmbH, Bremen, Germany) uniformly coating the sections with matrix crystals. As an organic matrix substance for the MALDI measurements 10 mg/mL sinapinic acid (Sigma-Aldrich, Taufkirchen, Germany) dissolved in water/acetonitrile 40:60 (v/v) with 0.2 % trifluoroacetic acid (TFA, Applied Biosystems, Darmstadt, Germany) were used.

Proteomic analysis by MALDI IMS

All imaging mass spectrometric measurements were performed on an Ultraflex III MALDI-TOF/TOF mass spectrometer and controlled by the software FlexControl 3.0 and FlexImaging 2.1 (Bruker Daltonik GmbH, Bremen, Germany). During the MALDI IMS measuring process a pulsed UV laser desorbed and ionized analytes to obtain mass signals directly from these tissue sections. At constant laser power and within the mass range of 2.5 to 20 kDa the mass-to-charge ratios (*m/z*) of the singly protonated molecular ions ($[M+H]^+$) were detected using a time-of-flight analyzer (TOF) in positive linear mode with a sampling rate of 0.1 GS/s. The software FlexImaging generates a mass spectrometric image with the values of the specific molecular masses and the peak intensities that were taken of thousands of spectra from a spot raster over the tissue. The spectra are based on an automatic accumulation of 200 randomly positioned laser shots per each raster position. The spot raster was set with a lateral resolution of 50 to 70 µm, depending on the size of the tissue section and the obtained mass spectrometric data, in order to find a good compromise between the resulting file size and the best possible spatial resolution power. This allowed the acquisition of high-resolution MS images. For calibration of the spectra the protein calibration standard I (Bruker Daltonik GmbH, Bremen, Germany) with known *m/z* values (insulin, ubiquitin I, cytochrome C und myoglobin) compounded with the matrix was coated on the same ITO coated glass slide next to the tissue section. The protein standard covered a mass range of 5 to 17.5 kDa. On the basis of the determined *m/z* values of the protein standard the *m/z* axis was calibrated. After the MALDI IMS measurements the measured tissue sections remained intact and matrix on the surface was rinsed off by 70 % ethanol prior to staining with hematoxylin and eosin (H&E). A microscope image of the H&E-stained tissue sections was taken with the Digital Slide System Pannoramic DESK (3DHISTECH Ltd., Budapest, Hungary). The captured digital

Table 1 Pathological characteristics of HCCs from the training and from the validation cohort.

case no.	gender	etiology	grade of differentiation	pT classification	pN classification	invasion into lymphatic vessels	vascular invasion
training set							
1	M	alcohol	moderate	3	0	0	0
2	M	alcohol	well	2	0	0	1
3	F	X	moderate	3	0	0	0
4	F	X	poor	3	0	0	1
5	M	X	moderate	2	0	0	0
6	M	X	well-moderate	2	0	0	X
7	F	X	moderate	3	0	1	0
8	M	X	moderate	3	1	1	0
9	M	X	moderate	3	X	0	0
10	M	NASH	moderate	1	0	0	0
11	F	X	moderate	1	0	0	0
12	M	X	poor	2	0	0	1
13	F	X	moderate	1	X	0	0
14	M	X	moderate	2	X	0	1
15	M	X	poor	3	X	X	1
validation set							
16	M	X	well	X	X	X	X
17	M	X	well	3	0	0	1
18	M	X	moderate	3	X	0	0
19	M	alcohol	poor	4	0	X	1
20	M	X	X	2	X	X	1
21	M	alcohol	moderate	3	0	0	1
22	M	alcohol	moderate	3	0	0	1
23	M	X	moderate	2	0	0	0
24	F	NASH	moderate	1	0	0	0
25	M	NASH	moderate	2	0	0	0

X = unknown; NASH, non-alcoholic steatohepatitis.

image was co-registered with the generated mass spectrometric image to correlate MALDI IMS data with the histological architecture of the tissue section. The complete tissue samples of patients with pathologically confirmed HCC were independently measured by MALDI IMS and randomly divided into a training set and a validation set.

Data analysis

In order to identify differentiating *m/z* values in malignant and non-malignant tissue components, the entire data set was statistically analyzed by the ClinProTools 2.1 software (Bruker Daltonik GmbH, Bremen, Germany). For better comparability automated preprocessing including baseline subtraction on spectra using the convex hull baseline algorithm, normalization to their own total ion count, recalibration of spectra, total average spectrum calculation, peak detection on the total average spectrum, peak calculation in the individual spectra and normalization of peak areas for model generation was performed. Diagnosis of HCC was verified by a second pathologist. Based on the histology of each individual tissue section, regions of interest (ROIs) corresponding to the tumor cell populations and to the non-neoplastic tissue components were defined using the FlexImaging 2.1 software (Bruker Daltonik GmbH, Bremen, Germany). The exported spectra of these ROIs were subdivided in different classes. The statistical comparison between these classes was performed using the Mann-Whitney U test with correction for multiple testing. P-values of < 0.05 were considered to be statistically significant. The diagnostic accuracy of the detected *m/z* values was determined using the area under the curve of the receiver operating characteristic (ROC) curve. For detecting and localizing specific proteomic signatures in the training set, we analyzed areas of tumor cells (n = 15) and adjacent fibrosis in the tumor tissue as well as fibrosis in the tissue sections, where no tumor cells were located (n = 11). The differentially expressed peaks with the highest statistical significance and diagnostic accuracy, which were detected in the tissues of the training set, were used to generate a genetic algorithm-based classification model. The proteomic classification model was applied to an independent validation set of tissue samples with malignant lesions of HCC (n = 10) and non-malignant regenerative nodules (n = 8).

Results

In the training set comparative analysis of 116 detected mass peaks in the tissue sections, four peaks at an average of *m/z* 6274, *m/z* 6647, *m/z* 6222 and *m/z* 6853 demonstrated a significantly higher expression in HCC lesions compared to non-malignant fibrotic areas (**Fig. 1**). The accuracy and significance levels of each peak to predict HCC individually are displayed in **Table 2**. The four peak-associated proteins were overexpressed in all 15 cases of the HCC tissue specimens in the training set (**Fig. 3**). Although the correlation of mass spectra with tissue sections identified in some cases *m/z* species that are lower expressed in HCC (e. g., m/z 3463 shown in **Fig. 2**) a statistical significance was not reached for these *m/z* species.

To determine a specific protein signature and to classify primary HCCs by MALDI IMS, the most significantly distinguishing peaks, as determined by the p values and the highest value of the AUROC at an average of *m/z* 6274, *m/z* 6647 and *m/z* 6222, were included

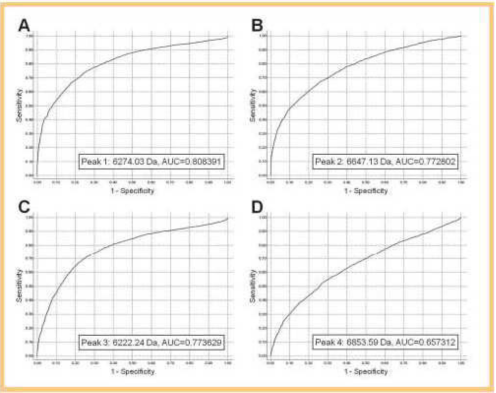


Fig. 1 Receiver operating characteristic (ROC) curve of the four significantly differentially expressed masses [training set] resulted in areas under the curve of 0.81 for m/z 6274 **A**, of 0.77 for m/z 6647 **B**, of 0.77 for m/z 6222 **C** and of 0.66 for m/z 6853 **D**. Da = Dalton.

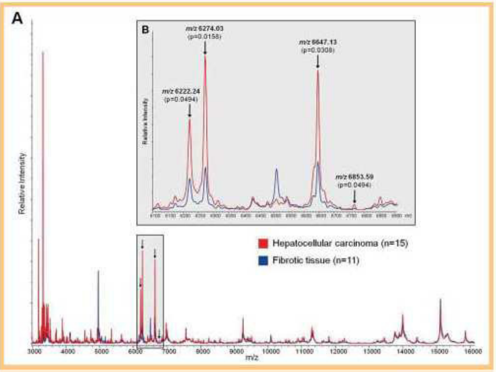


Fig. 3 **A** Average mass spectra in the mass range of 3 to 16 kDa obtained from the HCC (red) and fibrotic (blue) tissue regions of the training set. **B** Magnification of the mass range of 6100 to 6900 Da including the four significantly differentially expressed masses of m/z 6274, m/z 6647, m/z 6222, and m/z 6853, which are indicated by arrows.

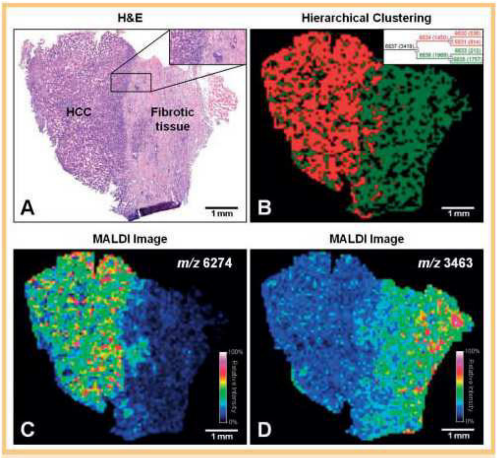


Fig. 2 MALDI IMS analysis of an individual HCC tissue sample (scale bars = 1 mm). **A** HE-stained histological image of the intact HCC tissue section with region of HCC cancer cells (left) and fibrosis (right) subsequent to the MALDI measurement. **B, D** MALDI IMS data of the very same HCC tissue section (lateral resolution of the mass spectrometric measurement: 50 µm). **B** Hierarchical clustering of the mass spectrometric dataset (at the top right corner displayed as a dendogram). Two different clusters containing spectra of the hepatocellular cancer cells (red) and fibrotic liver tissue (green) are represented. **C, D** Spatial distribution of two cell-type-specific m/z values in a ion-density map, which is a color-coded visualization of the relative intensity of the selected m/z value in two dimensions to represent graphically the distribution and level of expression of the underlying protein across the tissue. The color-intensity scale specifies the relative intensity of the corresponding m/z value from 0 % (black) to 100 % (white). **C** Heat map of m/z 6274 as a specific m/z value for cancer cells. **D** Heat map of m/z 3463 as a specific m/z value for the fibrotic liver tissue. Scale bars = 1 mm.

Table 2 Significantly different peaks of malignant (HCC) and non-malignant (fibrotic tissue) areas in the training set indicated by p-value and AUROC.

peak no.	MW (m/z)	p-value	AUROC
peak 1:	*m/z* 6274.03	p = 0.0158	0.808391
peak 2:	*m/z* 6647.13	p = 0.0308	0.772802
peak 3:	*m/z* 6222.24	p = 0.0494	0.773629
peak 4:	*m/z* 6853.59	p = 0.0494	0.650262

Peaks 1 – 3 were included in the genetic algorithm-based classification model based on their p-values and highest value of the AUROC.
MW = molecular weight; *m/z* = mass to charge ratio; AUROC = area under the curve of the receiver operating characteristic,

in a genetic algorithm-based classification model. The resulting classification model was applied to an independent validation set of tissue samples with lesions of HCC (n = 10) and regenerative nodules (n = 8). A cut-off value for a correct classification was defined as at least 60 % of the correctly classified HCC spectra. Based on these characteristics the classification algorithm classified one case of the HCC as false-negative in the non-malignant group (case number 19) and one case of the regenerative nodules as false-positive in the malignant group (case number 27). Thus, the proteomic classification model allowed a reliable prediction of lesions of HCC and of regenerative nodules with accuracy, sensitivity, specificity, positive predictive value and negative predictive value of 90 % each (**▶ Table 3**).

Discussion

In this study the histology-directed and label-free MALDI IMS technique was applied to discriminate HCC in human hepatic tissue. We were able to identify a specific protein signature of malignant lesions in the liver tissue. The combination of three differentially expressed masses with an average of *m/z* 6274, *m/z* 6647 and *m/z* 6222 overexpressed in the tumor regions enabled us to classify tissue into malignant and non-malignant components. The used genetic algorithm for classification belongs to the evolu-

Table 3 Predictive accuracy of the classification model for malignant (HCC) and non-malignant (regenerative nodules) lesions of the validation set.

case no.	spectra in total	excluded not recalibratable/null spectrum excluded	classified spectra in total	classified in class 1 (HCC)		classified in class 2 (fibrosis)		classified as
1 reg. nod.	771	27	744	431	(57.93 %)	313	(42.07 %)	class 2 (fibrosis)
5 reg. nod.	1216	11	1205	103	(8.55 %)	1102	(91.45 %)	class 2 (fibrosis)
8 reg. nod.	3656	185	3471	1389	(40.02 %)	2082	(59.98 %)	class 2 (fibrosis)
20 reg. nod.	90	2	88	22	(25.00 %)	66	(75.00 %)	class 2 (fibrosis)
26 reg. nod.	324	7	317	2	(0.63 %)	315	(99.37 %)	class 2 (fibrosis)
27 reg. nod.	2094	28	2066	1601	(77.49 %)	465	(22.51 %)	class 1 (HCC)
30 reg. nod.	75	1	74	0	(0.00 %)	74	(100.00 %)	class 2 (fibrosis)
31 reg. nod.	152	103	49	0	(0.00 %)	49	(100.00 %)	class 2 (fibrosis)
16 HCC	2638	927	1711	1562	(91.29 %)	149	(8.71 %)	class 1 (HCC)
17 HCC	1198	430	768	766	(99.74 %)	2	(0.26 %)	class 1 (HCC)
18 HCC	1877	295	1582	1441	(91.09 %)	141	(8.91 %)	class 1 (HCC)
19 HCC	855	13	842	293	(34.80 %)	549	(65.20 %)	class 2 (fibrosis)
20 HCC	149	0	149	131	(87.92 %)	18	(12.08 %)	class 1 (HCC)
21 HCC	3026	542	2484	1587	(63.89 %)	897	(36.11 %)	class 1 (HCC)
22 HCC	4719	13	4706	3319	(70.53 %)	1387	(29.47 %)	class 1 (HCC)
23 HCC	3666	524	3142	2288	(72.82 %)	854	(27.18 %)	class 1 (HCC)
24 HCC	3554	1003	2551	2174	(85.22 %)	377	(14.78 %)	class 1 (HCC)
25 HCC	1829	193	1636	1533	(93.70 %)	103	(6.30 %)	class 1 (HCC)
classification model (GA)								
accuracy			90 %					
sensitivity			90 %					
specificity			90 %					
positive predictive value			90 %					
negative predictive value			90 %					

reg. nod. = regenerative nodule; HCC = hepatocellular carcinoma; GA = genetic algorithm.

tionary algorithms which are optimization and search techniques based on the biological evolution. Starting from a randomly chosen set of initial solutions the genetic algorithm approaches step by step the optimal solution of the problem by repeatedly changing, combining and selecting of feasible solutions. Genetic algorithms are especially suitable when the structure of the problem is not well known or the number of possible solutions is very large. Beside several other classification algorithms, the genetic algorithm has already been successfully applied in various MS [32] and MALDI IMS studies [33].

The proteins found with MALDI IMS could serve as a basis for further investigations of possible potential biomarkers in tumor tissue. The results exhibited in this explorative study underline the great potential of the MALDI IMS technology for the diagnostic of HCC.

In spite of the convincing data, several limitations of this study should be mentioned. On the one hand, the number of the observed peaks, which were able to distinguish between the class of tumor tissues and the non-neoplastic tissues, of only four *m/z* values is quite small. On the other hand, due to the necessity of using frozen tissues with predefined compartments of tumor cells, the size of the sample collective of the study is also limited and, furthermore, the etiology of many samples is unknown. In view of the fact that many well-known protein biomarkers, which are applied in clinical practice as well as in clinical research in order to confirm the diagnosis, have a large molecular mass of more than 25 kDa (alpha-fetoprotein ≈ 67 kDa [34], GPC3 ≈ 70 kDa [35], HSP-70 ≈ 70 kDa [36], GS ≈ 42 kDa [37]), the observed typical mass range of this MS technology (between 2.4 and 25 kDa) is relatively restricted. Therefore, with MALDI IMS it is not possible to detect directly these known tumor markers. Furthermore, it is possible that because of the still relatively limited lateral resolution of the MALDI IMS technique of > 20 µm provided by commercially available devices, not every laser shot hits a tumor cell and, therefore, leads to restrictions of the statistical evaluation, especially in the case of very small sample material.

The detected proteins might represent unknown proteins with lower molecular mass or fragments of proteins with higher molecular mass, which are co-expressed in HCC and which could play a role in reaction mechanisms or in signalling pathways of HCC, respectively. Changes in the proteolytic activity of a variety of different proteases during tumor growth and spread, angiogenesis and apoptotic processes in HCC and in the adjacent tissue could be a possible cause for the formation of peptides, proteins or protein fragments. Such peptides/proteins or their fragments may correlate with the tumor tissue.

Our results demonstrate that with small amounts of fresh-frozen tissues, which can be obtained easily in the clinical practice, the resulting protein profiles from untreated tissue samples can be used to determine tissue-specific pattern of HCC. This pattern can be a support for the pathologist to reinforce the histopathological diagnosis. Since even small tissue samples, such as biopsy tissues or tissue microarrays, can be used in MALDI IMS [17, 21], the resulting protein profile could also be detected in suspicious tissues removed by ultrasound-guided needle biopsy to lead patients with small tumors or unclear imaging findings to an early therapy.

In a recently published study Le Faouder et al. identified a set of proteins/peptides by MALDI IMS that delineates HCC from adjacent cirrhotic tissue. Furthermore, by applying a generated classification model using a support vector machine one is able to discriminate between HCC and the peritumoral area in tissue sections of an independent validation set. This analysis detected 13 *m/z* values which were differentially expressed in regions of HCC and of cirrhosis in a training and in an independent validation cohort. The most discriminating peak (*m/z* 8565) overexpressed in HCC was identified as a monomeric form of ubiquitin, Ubi (1 – 74) [38]. In a continuative study, which was published recently, the authors showed that kallikrein-related peptidase 6 (KLK6) catalyzed the production of Ubi (1 – 74) from monomeric ubiquitin. Furthermore, they demonstrated that KLK6 was induced *de novo* in cirrhosis and found increased in HCC simultaneously with an accumulation of Ubi (1 – 74). The authors investigated *in vitro* the possible consequences of Ubi (1 – 74) accumulation, and demonstrated that Ubi (1 – 74) interferes with the normal ubiquitination machinery which is a complex network of proteins that control numerous cellular functions. Deregulation of this network leads to various human diseases, including development of many tumor types [39]. In another work of this research group the tissue proteome of HCC samples with and without microvascular invasion was compared using a MALDI IMS-based proteomic approach. In a comparative analysis of the acquired mass spectra 30 differential protein peaks were detected, among which 28 were more strongly expressed in HCC with microvascular invasion than in HCC without microvascular invasion. Two peaks (*m/z* 3790 and 5654) associated with microvascular invasion in HCC were identified as modified forms of histone H4. The results were validated in an independent series [40].

Han et al. also generated molecular profiles directly from the liver tissue in a limited number of HBV-related HCC patients using MALDI IMS. However, none of the obtained masses in this study was found in the data sets of Le Faouder et al. or in our comparative analysis. This seems to be due to the fact that in the study of Han et al. tissue sections of HCC were compared with non-tumoral liver sections that in most cases did not show cirrhosis [41].

Although our data set showed no significant differences in the expression of *m/z* 8565 in lesions of HCC, nevertheless two of the differentially expressed *m/z* values found in the study of Le Faouder et al. were identical to the two most significant peaks of our spectral data. These results support the hypothesis that the MALDI IMS method provides high accuracy and reproducibility to identify lesions of HCC in non-malignant liver tissue. Given a further validation in larger cohorts, the MALDI IMS technology could have a major impact on the clinical treatment of patients with HCC.

However, one must take into account that one has just started to apply the MS technology as an imaging technique to examine liver tissue sections in clinical research. Due to the rapid progress of this technology concerning mass accuracy, sensitivity, speed and lateral resolution as well as in bioinformatic aspects, it is to be expected that its application will improve considerably the diagnostic pathology in the near future. Through further analysis and identification of such differentially expressed *m/z* species one can possibly make also statements about the carcinogenesis and the aggressive characteristics of HCC. Besides the identification of new diagnostic markers, this method could thus assist in searching for new therapeutically valuable target molecules. Even if the molecular identity of the proteins of this new HCC-associated protein signature is currently unknown, the data can be validated through its identification and can provide an indication of their biological effect.

In summary, it was shown that the tissue proteome analysis by the use of MALDI IMS technology allows distinguishing between HCC and non-neoplastic liver tissue. Subject to a further validation of these results by independent and larger cohorts of patients and protein identification, the statistical significance of the results confirm that MALDI IMS proteome analysis of tissue can be a useful technique for molecular diagnostics in clinical applications.

Affiliations

[1] Department of Internal Medicine IV, Gastroenterology, Hepatology and Infectious Diseases, Jena University Hospital, Jena
[2] Experimental Transplantation Surgery, Department of General, Visceral and Vascular Surgery, Jena University Hospital, Jena
[3] Institute of Organic Chemistry and Macromolecular Chemistry, Friedrich-Schiller-University Jena
[4] Jena Center for Soft Matter (JCSM), Friedrich-Schiller-University Jena
[5] Department of General, Visceral and Vascular Surgery, Jena University Hospital, Jena
[6] Institute of Pathology, Jena University Hospital, Jena

Acknowledgements

Financial support from Interdisciplinary Center for Clinical Research (IZKF) of the University Medical Center Jena is gratefully acknowledged. The authors thank the Thüringer Ministerium für Bildung, Wissenschaft und Kultur (grant no. B515 – 07008) for the financial support of this study. They also thank Bruker Daltonik GmbH for their cooperation, Tony Bruns for a lot of helpful discussions and Elke Oswald for the very valuable technical assistance.

References

1 *Parkin DM, Bray F, Ferlay J et al.* Global Cancer Statistics, 2002. CA Cancer J Clin 2005; 55: 74 – 108. DOI: 10.3322/canjclin.55.2.74
2 *Llovet JM, Burroughs A, Bruix J.* Hepatocellular carcinoma. The Lancet 2003; 362: 1907 – 1917. DOI: 10.1016/S0140-6736(03)14964-1
3 *Poon D, Anderson BO, Chen LT et al.* Management of hepatocellular carcinoma in Asia: consensus statement from the Asian Oncology Summit 2009. Lancet Oncol 2009; 10: 1111 – 1118. DOI: 10.1016/S1470-2045(09)70241-4
4 *Sherman M.* Epidemiology of hepatocellular carcinoma. Oncology 2010; 78 (Suppl 1): 7 – 10. DOI: 10.1159/000315223
5 Neoplasia ICG for H. Pathologic diagnosis of early hepatocellular carcinoma: A report of the international consensus group for hepatocellular neoplasia. Hepatology 2009; 49: 658 – 664. DOI: 10.1002/hep.22709
6 *Di Tommaso L, Destro A, Seok JY et al.* The application of markers (HSP70 GPC3 and GS) in liver biopsies is useful for detection of hepatocellular carcinoma. J Hepatol 2009; 50: 746 – 754. DOI: 10.1016/j.jhep.2008.11.014
7 *Bruix J, Sherman M.* Management of hepatocellular carcinoma: An update. Hepatol Baltim Md 2011; 53: 1020 – 1022. DOI: 10.1002/hep.24199
8 *Stoeckli M, Chaurand P, Hallahan DE et al.* Imaging mass spectrometry: a new technology for the analysis of protein expression in mammalian tissues. Nat Med 2001; 7: 493 – 496. DOI: 10.1038/86573
9 *Schwartz SA, Weil RJ, Thompson RC et al.* proteomic-based prognosis of brain tumor patients using direct-tissue matrix-assisted laser desorption ionization mass spectrometry. Cancer Res 2005; 65: 7674 – 7681. DOI: 10.1158/0008-5472.CAN-04-3016
10 *Chaurand P, Schwartz SA, Caprioli RM.* Assessing protein patterns in disease using imaging mass spectrometry. J Proteome Res 2004; 3: 245 – 252
11 *Chaurand P, Sanders ME, Jensen RA et al.* Proteomics in diagnostic pathology: profiling and imaging proteins directly in tissue sections. Am J Pathol 2004; 165: 1057 – 1068. DOI: 10.1016/S0002-9440(10)63367-6

12 *Rauser S, Marquardt C, Balluff B et al.* Classification of HER2 receptor status in breast cancer tissues by MALDI imaging mass spectrometry. J Proteome Res 2010; 9: 1854–1863. DOI: 10.1021/pr901008d

13 *Reyzer ML, Caldwell RL, Dugger TC et al.* Early changes in protein expression detected by mass spectrometry predict tumor response to molecular therapeutics. Cancer Res 2004; 64: 9093–9100. DOI: 10.1158/0008-5472.CAN-04-2231

14 *Bauer JA, Chakravarthy AB, Rosenbluth JM et al.* Identification of markers of taxane sensitivity using proteomic and genomic analyses of breast tumors from patients receiving neoadjuvant paclitaxel and radiation. Clin Cancer Res 2010; 16: 681–690. DOI: 10.1158/1078-0432.CCR-09-1091

15 *Seeley EH, Caprioli RM.* Imaging mass spectrometry: Towards clinical diagnostics. PROTEOMICS – Clin Appl 2008; 2: 1435–1443. DOI: 10.1002/prca.200800013

16 *Seeley EH, Caprioli RM.* Molecular imaging of proteins in tissues by mass spectrometry. Proc Natl Acad Sci USA 2008; 105: 18126–18131. DOI: 10.1073/pnas.0801374105

17 *Kim HK, Reyzer ML, Choi IJ et al.* Gastric cancer-specific protein profile identified using endoscopic biopsy samples via MALDI mass spectrometry. J Proteome Res 2010; 9: 4123–4130. DOI: 10.1021/pr100302b

18 *Morita Y, Ikegami K, Goto-Inoue N et al.* Imaging mass spectrometry of gastric carcinoma in formalin-fixed paraffin-embedded tissue microarray. Cancer Sci 2010; 101: 267–273. DOI: 10.1111/j.1349-7006.2009.01384.x

19 *Deininger SO, Ebert MP, Fütterer A et al.* MALDI imaging combined with hierarchical clustering as a new tool for the interpretation of complex human cancers. J Proteome Res 2008; 7: 5230–5236. DOI: 10.1021/pr8005777

20 *Balluff B, Elsner M, Kowarsch A et al.* Classification of HER2/neu status in gastric cancer using a breast-cancer derived proteome classifier. J Proteome Res 2010; 9: 6317–6322. DOI: 10.1021/pr100573s

21 *Groseclose MR, Massion PP, Chaurand P et al.* High-throughput proteomic analysis of formalin-fixed paraffin-embedded tissue microarrays using MALDI imaging mass spectrometry. Proteomics 2008; 8: 3715–3724. DOI: 10.1002/pmic.200800495

22 *Yanagisawa K, Shyr Y, Xu BJ et al.* Proteomic patterns of tumour subsets in non-small-cell lung cancer. Lancet 2003; 362: 433–439. DOI: 10.1016/S0140-6736(03)14068-8

23 *Patel SA, Barnes A, Loftus N et al.* Imaging mass spectrometry using chemical inkjet printing reveals differential protein expression in human oral squamous cell carcinoma. The Analyst 2009; 134: 301–307. DOI: 10.1039/b812533c

24 *Lemaire R, Menguellet SA, Stauber J et al.* Specific MALDI imaging and profiling for biomarker hunting and validation: fragment of the 11S proteasome activator complex, Reg alpha fragment, is a new potential ovary cancer biomarker. J Proteome Res 2007; 6: 4127–4134. DOI: 10.1021/pr0702722

25 *Kang S, Shim HS, Lee JS et al.* Molecular proteomics imaging of tumor interfaces by mass spectrometry. J Proteome Res 2010; 9: 1157–1164. DOI: 10.1021/pr900666q

26 *Djidja M-C, Claude E, Snel MF et al.* MALDI-ion mobility separation-mass spectrometry imaging of glucose-regulated protein 78 kDa (Grp78) in human formalin-fixed, paraffin-embedded pancreatic adenocarcinoma tissue sections. J Proteome Res 2009; 8: 4876–4884. DOI: 10.1021/pr900522m

27 *Walch A, Rauser S, Deininger S-O et al.* MALDI imaging mass spectrometry for direct tissue analysis: a new frontier for molecular histology. Histochem Cell Biol 2008; 130: 421–434. DOI: 10.1007/s00418-008-0469-9

28 *Schwamborn K, Krieg RC, Reska M et al.* Identifying prostate carcinoma by MALDI-Imaging. Int J Mol Med 2007; 20: 155–159

29 *Cazares LH, Troyer D, Mendrinos S et al.* Imaging mass spectrometry of a specific fragment of mitogen-activated protein kinase/extracellular signal-regulated kinase kinase 2 discriminates cancer from uninvolved prostate tissue. Clin Cancer Res 2009; 15: 5541–5551. DOI: 10.1158/1078-0432.CCR-08-2892

30 *Oppenheimer SR, Mi D, Sanders ME et al.* Molecular analysis of tumor margins by MALDI mass spectrometry in renal carcinoma. J Proteome Res 2010; 9: 2182–2190. DOI: 10.1021/pr900936z

31 *Meding S, Nitsche U, Balluff B et al.* Tumor classification of six common cancer types based on proteomic profiling by MALDI imaging. J Proteome Res 2012; 11: 1996–2003. DOI: 10.1021/pr200784p

32 *Saeys Y, Inza I, Larrañaga P.* A review of feature selection techniques in bioinformatics. Bioinformatics 2007; 23: 2507–2517. DOI: 10.1093/bioinformatics/btm344

33 *Balluff B, Schöne C, Höfler H et al.* MALDI imaging mass spectrometry for direct tissue analysis: technological advancements and recent applications. Histochem Cell Biol 2011; 136: 227–244. DOI: 10.1007/s00418-011-0843-x

34 *Pucci P, Siciliano R, Malorni A et al.* Human alpha-fetoprotein primary structure: a mass spectrometric study. Biochemistry (Mosc) 1991; 30: 5061–5066

35 *Ho M, Kim H.* Glypican-3: a new target for cancer immunotherapy. Eur J Cancer Oxf Engl 2011; 47: 333–338. DOI: 10.1016/j.ejca.2010.10.024

36 *Sherman M, Multhoff G.* Heat shock proteins in cancer. Ann N Y Acad Sci 2007; 1113: 192–201. DOI: 10.1196/annals.1391.030

37 *Osada T, Sakamoto M, Nagawa H et al.* Acquisition of glutamine synthetase expression in human hepatocarcinogenesis: relation to disease recurrence and possible regulation by ubiquitin-dependent proteolysis. Cancer 1999; 85: 819–831

38 *Le Faouder J, Laouirem S, Chapelle M et al.* Imaging mass spectrometry provides fingerprints for distinguishing hepatocellular carcinoma from cirrhosis. J Proteome Res 2011; 10: 3755–3765. DOI: 10.1021/pr200372p

39 *Laouirem S, Le Faouder J, Alexandrov T et al.* Progression from cirrhosis to cancer is associated with early ubiquitin post-translational modifications: identification of new biomarkers of cirrhosis at risk of malignancy. J Pathol 2014; DOI: 10.1002/path.4398

40 *Poté N, Alexandrov T, Le Faouder J et al.* Imaging mass spectrometry reveals modified forms of histone H4 as new biomarkers of microvascular invasion in hepatocellular carcinomas. Hepatol Baltim Md 2013; 58: 983–994. DOI: 10.1002/hep.26433

41 *Han EC, Lee YS, Liao WS et al.* Direct tissue analysis by MALDI-TOF mass spectrometry in human hepatocellular carcinoma. Clin Chim Acta 2011; 412: 230–239. DOI: 10.1016/j.cca.2010.09.021

4. Discussion

Patients with developed HCC mostly have no symptoms besides those related to their underlying chronic liver disease and cirrhosis (Sugano et al., 1994). In general, HCC mostly discovered while routine screening or because of dominant symptoms associated with tumor size or location. HCC tumors are presented as a single lesion or as a diffuse mass, what can be difficult to distinguish from the surrounding cirrhotic tissue and regenerative nodules by available imaging studies (Kuntz and Kuntz, 2008). Nowadays, only aggressive measures such as surgical resection, ablative therapy or liver transplantation provide promising survival prognosis for the patients with HCC. Against that background, novel molecular imaging techniques should be considered in order to diagnose and predict the extremely malignant and rapidly progressing liver tumor at the early stage.

4.1 Application of Raman spectroscopy for liver cancer investigations

Different optical techniques such as Raman, Surface-enhanced Raman (SERS), Coherent anti-Stokes Raman spectroscopy (CARS) and Laser tweezers Raman (LTRS) spectroscopies as well as Raman spectroscopy in combination with Laser-induced fluorescence (LIF) have been applied for liver cancer diagnostics (Guo et al., 2009; Hawi et al., 1996; Li et al., 2004; Santiago-Cordoba et al., 2011; Taleb et al., 2013; Wang et al., 2009; Wu et al., 2009; Xiong et al., 2012).

In 1996 Hawi *et al.* have published the first investigations of malignant and normal hepatocytes by single point Raman spectroscopy in culture and in tissue samples (Hawi et al., 1996). In this study they were able to detect spectral differences in single Raman bands and showed the future potential of this method. Later, similar work was performed by Guo *et al.* where they applied a linear discriminant analysis for identification of malignant hepatocytes (Guo et al., 2009). First work with serum samples from patients with HCC and a control group were investigated by Taleb *et al.* applying micro-Raman spectroscopy in 2013. Here principal component analysis (PCA) allowed separating patients into two groups (with and without HCC). In contrast, SVM classification model using leave-one-out cross validation algorithm provided an accuracy from 84.5 to 90.2 % for dried serum drops and from 86 to 91.5 % for freeze-dried serum drops (Taleb et al., 2013). Another work with serum samples was published by Santiago-Cordoba *et al.* (Santiago-Cordoba et al., 2011). In that work SERS was applied for serum analysis of the HCC patients with HBV or HCV infections and provide fast (< 10 s) and reproducible (<5 %variation) detection. In addition, Raman spectra of oxyhemoglobin of liver cancer patients (n = 30) and healthy persons (n = 30) were measured by SERS that allowed to discriminate between two groups with an accuracy of 93.3 % (Xiong et al., 2012). Another type of Raman spectroscopy, CARS, was used for the quantitative assessment of fat in intact liver tissue reported by Wu *et al.* (Wu et al., 2009). Along with it a LTRS has been applied for identification of liver cancer cells, para-cancer cells and normal hepatocytes (Wang et al., 2009). In this study, intensity differences at 1070, 1266, 1445 cm^{-1} were detected in the Raman spectra of different cell types. The article published in 2004 by Li *et al.* showed that patients without liver cancer had three sharp Raman bands in their serum samples in contrast to patients with liver cancer. Also the pick shift in fluorescence, that were observed for healthy patients compared to one with liver fibrosis and cirrhosis, was detected by LIF (Li et al., 2004).

Despite of all above mentioned methods, our research group has applied Raman spectroscopic imaging as an innovative technology for diagnostics of liver cancer for the first time. Raman imaging combines the spatial resolution of microscopic imaging with highly specific spectroscopic information of classical Raman spectroscopy. In order to achieve this goal, different studies to investigate liver cancer on the cellular and tissue levels for patients with HCC were planned and achieved (see Chapter Aim of the work).

Identification and classification of HCC cells by Raman imaging spectroscopy

The aim of this study was to differentiate and classify by Raman imaging spectroscopy a hepatocellular carcinoma HepG2 cell line (the most common primary liver cancer) and an adenocarcinoma SK-Hep1 cell line (found in an ascites fluid of the patient with liver cancer). In contrast to previous research studies, the identification of the different cell types was based on spectral information collected from the whole cell with the step size of 1 µm rather than from single point measurements, published by Goa *et al.* and Hawi *et al.* (Guo et al., 2009; Hawi et al., 1996). This allowed us to gain more specific chemical information and identify subcellular components, including nucleic acids, proteins, carbohydrates and cytoplasmic lipids within each measured cell. Thus, differences in Raman spectral information for different cellular organelles were detected. The difference spectra represent significant spectral changes in the Raman bands from 2900 to 2850 and at 1655, 1440, 1304, 1266 and 1060 cm^{-1}. Further analysis of this data with HCA has divided HepG2 and SK-Hep1 cells into two major clusters without any misclassification. The application of SVM classification algorithm to the data set allowed us to identify previously unknown cells with an accuracy of 93 %. Precise analysis of Raman spectra showed that two cell lines differed mainly in specific spectral patterns of unsaturated fatty acids. Therefore, the use of SVM algorithms for each cellular compartment separately improved the accuracy of the classification model up to 96 % for cytoplasmic lipids compare to nucleus, cytoplasm or complete cell information with accuracies of 91 %, 87 % and 93 %, respectively.

As was found in our investigations, the lipid synthesis plays an important role in carcinogenesis in the liver. In 2005, high expression of lipogenic enzymes by mRNA in HCC liver tissue was described by Yahagi *et al.* (Yahagi et al., 2005). Another study showed upregulation of lipogenic enzymes, which are involved in the fatty acid synthesis and cholesterol biosynthesis as well as their upstream inductors in non-tumorous liver tissue in contrast to HCC. Furthermore, this study discovered clinical aggressiveness based on a correlation of higher expression of enzymes with an activation of a lipogenic signalling pathway (Calvisi et al., 2011).

The second part of this study was based on analysis of the proliferation behavior of HCC cells by Raman imaging spectroscopy. The main difference between actively dividing cells and cells in apoptosis was found to be in the higher amount of fatty acids in dividing cells. Similar results were achieved by Gao *et al.* based on single point Raman spectroscopy. He showed that the inhibition of the fatty acid synthesis causes a cell cycle arrest of HepG2 cells (Gao et al., 2006). The state of cell proliferation can be predicted by Raman spectral bands for HepG2 cells in the exponential and plateau phase with sensitivity of 100 % and specificity of 98 %.

Based on the results of this study, in our forthcoming investigations we would like to focus on the Raman based analysis of brush cytology acquired during endoscopic retrograde cholangiopancreatography for diagnostics of CCA.

Identification and classification of HCC tissue by Raman imaging spectroscopy

The meaning of diagnostics in the clinical routine nowadays includes laboratory analyses, imaging studies (CT and MRI) as well as cytological and histological investigations of tissue samples. The identification of measurable indicators (biomarkers) in biological specimens is crucial for the development of an effective diagnostic method to identify the underlying disease. Raman spectroscopy in combination with confocal microscopy allows the detection of various markers in the sample. This “fingerprint” information can be used for objective tumor differentiation and classification of different types of tissue.

Following the previous study about classification of different types of liver cancer cells and their proliferation states, the aim of this project was to investigate different regions of human HCC tissue by Raman imaging spectroscopy. Therefore we measured predefined malignant (HCC, n = 12) and non-malignant (fibrosis, n = 17) tissue regions for comparative analysis. Despite high similarities of the Raman spectral signature of biomolecules, the chemometric data analysis can be applied in order to identify, classify and predict HCC. Precise analysis of different spectra allows us to detect changes in lipids together with proteins in HCC tissues and collagen together with proteins in fibrotic regions. Applying the N-FINDER algorithm we obtained four independent spectral endmembers for proteins, triglycerides, collagen and cholesterol esters. Differences in the distribution of the molecules in the Raman maps showed high heterogeneity of HCC samples with a prevailing amount of lipids or proteins depending on the patient. Because of the complexity of the data set, a nonlinear RF was chosen to predict HCC. A 101x iterated sevenfold cross-validated model resulted in sensitivity of 76 % and specificity of 93 % (accuracy of 86 %). Therefore, we can conclude that based on Raman imaging spectroscopy combined with RF classification model 94 patients with HCC from 100 will be correctly identified as having cancer. And 77 healthy patients from

100 will be identified as healthy. Nevertheless, 23 patients with HCC will be incorrectly identified as healthy and 6 healthy patients will be misclassified as having cancer. The variable importance plot demonstrated most significant bands for classification at the wavenumbers 2856, 1298, 937, 920, 1439, 1080, 2895, 1023, 1063 and 719 cm^{-1} (ranked in order of importance). The correlation of the Raman peaks with reference data bases of Raman spectroscopy showed that four out of ten bands can be identified as palmitic acid in HCC tissue (Bonifacio and Sergo, 2010; Bonifacio et al., 2010; De Gelder et al., 2007; Köhler et al., 2009; Krafft et al., 2005). Furthermore, we applied the classifier to Raman imaging maps of tumor margins and the bordering fibrotic tissue to confirm the prediction of the developed classifier. Therefore an algorithm that provides prediction maps with color-intensity scales depicting the positive diagnose for HCC was generated. The predictive performance of the algorithm by its correlation with the corresponding tissue architecture on H&E stained regions were verified.

The results of this study are in a good agreement with our previous article where we identified and classified liver cancer cells and proliferation states by Raman imaging. The main differences that we found between HCC and adenocarcinoma cells were in the high amount of unsaturated fatty acids in HCC cells and especially during exponential growth phase. It is known that high lipid biosynthesis plays an important role in carcinogenesis of HCC. Overexpression of mRNA for fatty acid synthase, acetyl-CoA carboxylase and ATP citrate lyase was described by Yahagi *et al.* (Yahagi et al., 2005). In additional upregulation of lipogenic enzymes, which is involved in the fatty acid synthesis and cholesterol biosynthesis, was detected in HCC tissue. Later an activation of a lipogenic pathway (AKT-mTORC1-RPS6) was correlated with aggressiveness of HCC (Calvisi et al., 2011). Subsequently, work published by Budhu *et al.* showed increased level of palmitic acid in aggressive HCCs, that confirm our results about Raman signatures of palmitic acid to predict HCC.

Different types of lesions and surgical margin status of HCC was predicted by Raman imaging spectroscopy with 86 % of accuracy. As a non-destructive and label-free imaging technique this method could allow to perform *in vivo* diagnostics of HCC without taking biopsies in the future. Thereby, Raman imaging can provide a fast prediction of cancer or other diseases in the clinics.

Data analysis of Raman spectroscopic data

Analysis and interpretation of the spectral information are crucial steps in Raman imaging. Detected molecular changes in biological specimens require the application of statistical methods (e.g. preprocessing, multivariate imaging, classification) in order to extract diagnostic information. After the sample collections and preparation, followed by Raman measurements we obtain raw Raman spectra that should be preprocessed by cosmic spike removal, quality test, baseline corrections, normalization and spectra calibration, if necessary (Bielecki et al., 2012; Kloss et al., 2013; Neugebauer et al., 2010b). These steps are important for extracting undesirable fluorescence, burning artifacts, contaminants and cosmic ray features, as well as for the intensity correlation that allows to have comparable data sets with recognizable differences. As the next steps the supervised (classification methods) and unsupervised (clustering methods) algorithmic approaches can be applied to the data sets. The classification models, like SVM or RF, allow to classify and predict unknown sample sets. At the same time the cluster analysis, like HCA or N-FINDER, allow combining data sets into the groups (clusters or endmembers), which correspond to different cell or tissue types, or even

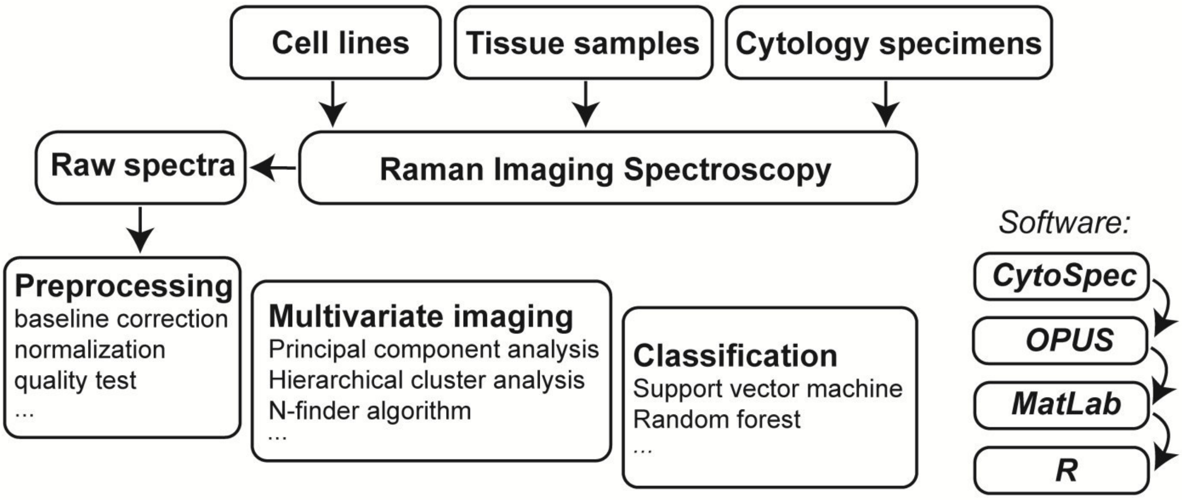


Figure 5. Schematic representation of processing pathway for Raman spectroscopic imaging of biological specimens introduced in the presented thesis.

different chemical components within the data set. This analysis can be performed by using programming languages Matlab and R with self-written scripts or available toolboxes as well as by applying such softwares like Cytospec or OPUS.

For all our experiments we used different chemometrical algorithms and developed a strategy for processing and interpretation of spectral information collected during Raman measurements, as can be seen in Figure 5.

4.2 Application of MALDI IMS for liver cancer investigations

During development of cancer, surrounding microenvironment and cells themselves synthetize molecules, such as proteins and peptides, in a different way as normal cells. These variations in types and concentrations of the molecules can be successfully analyzed by spectrometric methods and the patterns can help to identify specific biomarkers for clinical approval. MALDI IMS is considered to be a powerful method for molecular imaging with the great potential to be used as a diagnostic tool in studies of cancer (Norris and Caprioli, 2013).

Over recent years MALDI IMS has been used in investigation and prediction of different cancer types directly in tissue specimens (Chughtai et al., 2012; Dihazi et al., 2013; McDonnell et al., 2012). Furthermore, MALDI IMS showed a high potential in the field of gastroenterology and hepatology (Balluff et al., 2012). For instance, the study published by Cheng *et al.* showed that human neutrophil peptides 1-3 are synthetized in high amount in gastric cancer tissue and can be used for diagnostic and target tumor treatment (Cheng et al., 2012). MALDI IMS was applied for investigation of colon cancer cells and their metastasing into regional lymph nodes. In that study three novel biomarkers responsible for development of metastasis in lymph nodes were discovered (Meding et al., 2012).

The objective of our study was to investigate protein signatures in tissue samples in order to identify new specific molecular patterns for diagnostic and prediction of HCC by MALDI IMS. For this purpose we analyzed surgically resected liver specimens with predefined regions of HCC, regenerative nodules and fibrosis. From 116 detected mass values within the tissue samples we were able to identify four peaks (6274, 6647, 6222 and 6853 m/z) which displayed high expression profile in area of HCC in compare to non-malignant areas. In order to identify HCC based on three most significant protein peaks (6274, 6647 and 6222 m/z) detected by MALDI IMS, the genetic algorithm-based classification model was applied. This analysis allowed us to predict lesions of HCC and regenerative nodules with accuracy, sensitivity, and specificity of 90 %. What means, only 10 patients from 100 would be misclassified while diagnosis of HCC by MALDI IMS.

Recently published work by Le Faouder, showed similar investigations of cirrhotic liver tissue specimens with HCC by MALDI IMS (Le Faouder et al., 2011). Proteomic analysis was applied to frozen liver tissue (n = 30), where HCC regions were compared with cirrhosis in order to characterize new biomarkers. Using a SVM classification model based on 13 previously selected protein peaks sensitivity of 83.8 % and specificity of 86 % was achieved. The most distinguished protein peak was found at 8565 m/z and characterized as the monomeric ubiquitin (Ubi). Further investigations by this research group demonstrated role of kallikrein-related peptidase 6 that catalyzed the production of Ubi in translation from cirrhosis to HCC (Laouirem et al., 2014). Another study demonstrated the analysis of tissue proteome of HCC samples allowing to distinguish between samples with and without microvascular invasion (Poté et al., 2013). Acquired 28 peaks were highly expressed in HCC tissue. Furthermore, two of them (3790 and 5654 m/z) were detected as modified forms of histone H4 and associated with microvascular invasion in HCC. Moreover, application of MALDI IMS for analysis of liver tissue specimens allow classification of HCC tumors associated with HBV vs. non-tumor tissue regions (Han et al., 2011).

In spite of all previously presented data, in our study no significant correlations with results presented by Le Faouder *et al.* and Han *et al.* were found. Nevertheless, even if no information about overexpression of Ubi (8565 m/z) in HCC was obtain we can take into account that two of the proteins peaks, described by Le Faouder *et al.*, highly expressed in the HCC tissue were detected in cancer tissue regions measured by us as well. This proteomic analysis allows us not only to investigate the protein signatures, but as well to obtain their spatial information in HCC, fibrosis and regenerative nodules within liver tissue sections. In conclusion, by collecting mass spectrometry data from various liver tissues and cells new data base for a fast prediction of cancer or other diseases can be developed. Therefore in our study, we developed a new algorithm for classification and prediction of HCC vs. non-malignant tissue based on MALDI IMS. In our forthcoming studies we would like to focus on analysis of higher number of samples, as well as investigation of lipid distributions within the HCC tissue samples.

Conclusion

Molecular imaging techniques, such as Raman spectroscopy and MALDI IMS spectrometry, allow to investigate various of molecules and their spatial distribution within biological specimens. Raman spectroscopy is a biophotonic method for non-destructive and label-free visualization of vibrational and rotational states of molecules at submicron resolution. Raman spectra of biological samples includes the information about proteins, nucleic acids, lipids, carbohydrates and provide rapid and specific biochemical “fingerprint” for differentiation and classification of malignancies. In contrast, MALDI IMS is a powerful technique for analysis and identification of proteins or lipids in the biomedical specimens. This method allows to discover specific molecules as biomarkers for diagnostic of diseases. Furthermore, by applying the classification algorithms to the spectroscopic or spectrometric data, cancer-specific patterns can be discovered in order to develop independent prediction models. Therefore, by collecting molecular signatures from various liver tissues and cells by Raman and MALDI imaging, new data base for the fast prediction of cancer or other diseases can be developed.

In conclusion, we applied two molecular imaging methods for diagnostic of liver cancer on cellular and tissue levels and confirmed that Raman imaging spectroscopy and MALDI IMS in combination with multivariate data analysis are able to identify and predict liver cancer with high accuracy.

Bibliography

Aichler, M., and Walch, A. (2015). MALDI Imaging mass spectrometry: current frontiers and perspectives in pathology research and practice. Lab Invest *95*, 422–431.

Alexandrov, T., Becker, M., Guntinas-Lichius, O., Ernst, G., and von Eggeling, F. (2013). MALDI-imaging segmentation is a powerful tool for spatial functional proteomic analysis of human larynx carcinoma. J. Cancer Res. Clin. Oncol. *139*, 85–95.

Asghari-Khiavi, M., Vongsvivut, J., Perepichka, I., Mechler, A., Wood, B.R., McNaughton, D., and Bohle, D.S. (2011). Interaction of quinoline antimalarial drugs with ferriprotoporphyrin IX, a solid state spectroscopy study. J. Inorg. Biochem. *105*, 1662–1669.

Balluff, B., Rauser, S., Ebert, M.P., Siveke, J.T., Höfler, H., and Walch, A. (2012). Direct Molecular Tissue Analysis by MALDI Imaging Mass Spectrometry in the Field of Gastrointestinal Disease. Gastroenterology *143*, 544–549.e2.

Behne, T., and Copur, M.S. (2012). Biomarkers for Hepatocellular Carcinoma. International Journal of Hepatology *2012*.

Belghiti, J., Hiramatsu, K., Benoist, S., Massault, P., Sauvanet, A., and Farges, O. (2000). Seven hundred forty-seven hepatectomies in the 1990s: an update to evaluate the actual risk of liver resection. J. Am. Coll. Surg. *191*, 38–46.

Beljebbar, A., Bouché, O., Diébold, M.D., Guillou, P.J., Palot, J.P., Eudes, D., and Manfait, M. (2009). Identification of Raman spectroscopic markers for the characterization of normal and adenocarcinomatous colonic tissues. Crit. Rev. Oncol. Hematol. *72*, 255–264.

Bergholt, M.S., Zheng, W., Lin, K., Ho, K.Y., Teh, M., Yeoh, K.G., So, J.B.Y., and Huang, Z. (2010). Raman endoscopy for in vivo differentiation between benign and malignant ulcers in the stomach. Analyst *135*, 3162–3168.

Bergk, A., and Schott, E. (2011). HCC Prävention und Diagnostik bei Leberzirrhose. Hepatitis & More *1*, 48–52.

Bergner, N., Krafft, C., Geiger, K.D., Kirsch, M., Schackert, G., and Popp, J. (2012). Unsupervised unmixing of Raman microspectroscopic images for morphochemical analysis of non-dried brain tumor specimens. Anal Bioanal Chem *403*, 719–725.

Bielecki, C., Bocklitz, T.W., Schmitt, M., Krafft, C., Marquardt, C., Gharbi, A., Knösel, T., Stallmach, A., and Popp, J. (2012). Classification of inflammatory bowel diseases by means of Raman spectroscopic imaging of epithelium cells. J Biomed Opt *17*, 076030.

Bocklitz, T., Putsche, M., Stüber, C., Käs, J., Niendorf, A., Rösch, P., and Popp, J. (2009). A comprehensive study of classification methods for medical diagnosis. J. Raman Spectrosc. *40*, 1759–1765.

Bonifacio, A., and Sergo, V. (2010). Effects of sample orientation in Raman microspectroscopy of collagen fibers and their impact on the interpretation of the amide III band. Vibrational Spectroscopy *53*, 314–317.

Bonifacio, A., Beleites, C., Vittur, F., Marsich, E., Semeraro, S., Paoletti, S., and Sergo, V. (2010). Chemical imaging of articular cartilage sections with Raman mapping, employing uni- and multi-variate methods for data analysis. Analyst *135*, 3193–3204.

Bruix, J., Sherman, M., Llovet, J.M., Beaugrand, M., Lencioni, R., Burroughs, A.K., Christensen, E., Pagliaro, L., Colombo, M., Rodés, J., et al. (2001). Clinical management of hepatocellular carcinoma. Conclusions of the Barcelona-2000 EASL conference. European Association for the Study of the Liver. J. Hepatol. *35*, 421–430.

Calvisi, D.F., Wang, C., Ho, C., Ladu, S., Lee, S.A., Mattu, S., Destefanis, G., Delogu, S., Zimmermann, A., Ericsson, J., et al. (2011). Increased lipogenesis, induced by AKT-mTORC1-RPS6 signaling, promotes development of human hepatocellular carcinoma. Gastroenterology *140*, 1071–1083.

Chan, J.W., and Lieu, D.K. (2009). Label-free biochemical characterization of stem cells using vibrational spectroscopy. J Biophotonics *2*, 656–668.

Chan, J.W., Lieu, D.K., Huser, T., and Li, R.A. (2009). Label-free separation of human embryonic stem cells and their cardiac derivatives using Raman spectroscopy. Anal. Chem. *81*, 1324–1331.

Cheng, C.-C., Chang, J., Chen, L.-Y., Ho, A.-S., Huang, K.-J., Lee, S.-C., Mai, F.-D., and Chang, C.-C. (2012). Human neutrophil peptides 1-3 as gastric cancer tissue markers measured by MALDI-imaging mass spectrometry: implications for infiltrated neutrophils as a tumor target. Dis. Markers *32*, 21–31.

Chuang, S.-H., Li, J., Sun, X., Vadlamudi, A., Sun, B., Cazares, L., Nyalwidhe, J., Troyer, D., Semmes, J., and Mckenzie, F.D. (2012). Prostate Cancer Region Prediction by Fusing Results from MALDI Spectra-processing and Texture Analysis. Simulation *88*, 1247–1259.

Chughtai, S., Chughtai, K., Cillero-Pastor, B., Kiss, A., Agrawal, P., MacAleese, L., and Heeren, R.M.A. (2012). A multimodal mass spectrometry imaging approach for the study of musculoskeletal tissues. International Journal of Mass Spectrometry *325-327*, 150–160.

De Gelder, J., De Gussem, K., Vandenabeele, P., and Moens, L. (2007). Reference database of Raman spectra of biological molecules. J. Raman Spectrosc. *38*, 1133–1147.

Dieing, T., Hollricher, O., and Toporski, J. (2011). Confocal Raman Microscopy.

Dihazi, H., Bohrer, R., Jahn, O., Lenz, C., Majcherczyk, A., Schmidt, B., Urlaub, H., Valerius, O., and Asif, A.R. (2013). Mass spectrometry imaging: linking molecule profiles to tissue spatial distribution. Expert Rev Proteomics *10*, 17–20.

Dudley, E. (2014). MALDI Profiling and Applications in Medicine. In Advancements of Mass Spectrometry in Biomedical Research, A.G. Woods, and C.C. Darie, eds. (Springer International Publishing), pp. 33–58.

EASL, and European Organisation for Research and Treatment of Cancer (EORTC) (2012). EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma. J. Hepatol. *56*, 908–943.

Ellis, D.I., Cowcher, D.P., Ashton, L., O'Hagan, S., and Goodacre, R. (2013). Illuminating disease and enlightening biomedicine: Raman spectroscopy as a diagnostic tool. Analyst *138*, 3871–3884.

European Association For The Study Of The Liver, and European Organisation For Research And Treatment Of Cancer (2012). EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma. J. Hepatol. *56*, 908–943.

Forner, A., Vilana, R., Ayuso, C., Bianchi, L., Solé, M., Ayuso, J.R., Boix, L., Sala, M., Varela, M., Llovet, J.M., et al. (2008). Diagnosis of hepatic nodules 20 mm or smaller in cirrhosis: Prospective validation of the noninvasive diagnostic criteria for hepatocellular carcinoma. Hepatology *47*, 97–104.

Gao, Y., Lin, L.-P., Zhu, C.-H., Chen, Y., Hou, Y.-T., and Ding, J. (2006). Growth arrest induced by C75, A fatty acid synthase inhibitor, was partially modulated by p38 MAPK but not by p53 in human hepatocellular carcinoma. Cancer Biol. Ther. *5*, 978–985.

Gessel, M.M., Norris, J.L., and Caprioli, R.M. (2014). MALDI Imaging Mass Spectrometry: Spatial Molecular Analysis to Enable a New Age of Discovery. J Proteomics *0*, 71–82.

Ghita, A., Pascut, F.C., Mather, M., Sottile, V., and Notingher, I. (2012). Cytoplasmic RNA in undifferentiated neural stem cells: a potential label-free Raman spectral marker for assessing the undifferentiated status. Anal. Chem. *84*, 3155–3162.

Greten, T.F., Papendorf, F., Bleck, J.S., Kirchhoff, T., Wohlberedt, T., Kubicka, S., Klempnauer, J., Galanski, M., and Manns, M.P. (2005). Survival rate in patients with hepatocellular carcinoma: a retrospective analysis of 389 patients. Br. J. Cancer *92*, 1862–1868.

Groseclose, M.R., Massion, P.P., Chaurand, P., and Caprioli, R.M. (2008). High-throughput proteomic analysis of formalin-fixed paraffin-embedded tissue microarrays using MALDI imaging mass spectrometry. Proteomics *8*, 3715–3724.

Guarracino, M.R., Xanthopoulos, P., Pyrgiotakis, G., Tomaino, V., Moudgil, B.M., and Pardalos, P.M. (2011). Classification of cancer cell death with spectral dimensionality reduction and generalized eigenvalues. Artif Intell Med *53*, 119–125.

Guo, J., Du, B., Qian, M., Cai, W., Wang, Z., and Sun, Z. (2009). Raman spectroscopic identification of normal and malignant hepatocytes. Chin. Opt. Lett. *7*, 60–63.

Guze, K., Pawluk, H.C., Short, M., Zeng, H., Lorch, J., Norris, C., and Sonis, S. (2014). Pilot study: Raman spectroscopy in differentiating premalignant and malignant oral lesions from normal mucosa and benign lesions in humans. Head Neck.

Halperin, E.C., Brady, L.W., Perez, C.A., and Wazer, D.E. (2013). Perez & Brady's Principles and Practice of Radiation Oncology (Lippincott Williams & Wilkins).

Han, E.C., Lee, Y.-S., Liao, W.-S., Liu, Y.-C., Liao, H.-Y., and Jeng, L.-B. (2011). Direct tissue analysis by MALDI-TOF mass spectrometry in human hepatocellular carcinoma. Clin. Chim. Acta *412*, 230–239.

Harris, A.T., Garg, M., Yang, X.B., Fisher, S.E., Kirkham, J., Smith, D.A., Martin-Hirsch, D.P., and High, A.S. (2009). Raman spectroscopy and advanced mathematical modelling in the discrimination of human thyroid cell lines. Head & Neck Oncology *1*, 38.

Harz, M., Kiehntopf, M., Stöckel, S., Rösch, P., Deufel, T., and Popp, J. (2008). Analysis of single blood cells for CSF diagnostics via a combination of fluorescence staining and micro-Raman spectroscopy. Analyst *133*, 1416–1423.

Hawi, S.R., Campbell, W.B., Kajdacsy-Balla, A., Murphy, R., Adar, F., and Nithipatikom, K. (1996). Characterization of normal and malignant human hepatocytes by Raman microspectroscopy. Cancer Lett. *110*, 35–40.

Hedegaard, M., Krafft, C., Ditzel, H.J., Johansen, L.E., Hassing, S., and Popp, J. (2010). Discriminating isogenic cancer cells and identifying altered unsaturated fatty acid content as associated with metastasis status, using k-means clustering and partial least squares-discriminant analysis of Raman maps. Anal. Chem. *82*, 2797–2802.

Huser, T., Orme, C.A., Hollars, C.W., Corzett, M.H., and Balhorn, R. (2009). Raman spectroscopy of DNA packaging in individual human sperm cells distinguishes normal from abnormal cells. J Biophotonics *2*, 322–332.

International Agency for Research on Cancer (IARC), and World Health Organization (WHO) (2014). World cancer report 2014 (Lyon, France: International Agency for Research on Cancer).

Jess, P.R.T., Smith, D.D.W., Mazilu, M., Dholakia, K., Riches, A.C., and Herrington, C.S. (2007). Early detection of cervical neoplasia by Raman spectroscopy. Int. J. Cancer *121*, 2723–2728.

Kamemoto, L.E., Misra, A.K., Sharma, S.K., Goodman, M.T., Luk, H., Dykes, A.C., and Acosta, T. (2010). Near-infrared micro-Raman spectroscopy for in vitro detection of cervical cancer. Appl Spectrosc *64*, 255–261.

Kang, J.W., Lue, N., Kong, C.-R., Barman, I., Dingari, N.C., Goldfless, S.J., Niles, J.C., Dasari, R.R., and Feld, M.S. (2011). Combined confocal Raman and quantitative phase microscopy system for biomedical diagnosis. Biomed Opt Express *2*, 2484–2492.

Kloss, S., Kampe, B., Sachse, S., Rösch, P., Straube, E., Pfister, W., Kiehntopf, M., and Popp, J. (2013). Culture independent Raman spectroscopic identification of urinary tract infection pathogens: a proof of principle study. Anal. Chem. *85*, 9610–9616.

Köhler, M., Machill, S., Salzer, R., and Krafft, C. (2009). Characterization of lipid extracts from brain tissue and tumors using Raman spectroscopy and mass spectrometry. Anal Bioanal Chem *393*, 1513–1520.

Kourtis, N., and Tavernarakis, N. (2009). Autophagy and cell death in model organisms. Cell Death Differ. *16*, 21–30.

Krafft, C., Neudert, L., Simat, T., and Salzer, R. (2005). Near infrared Raman spectra of human brain lipids. Spectrochim Acta A Mol Biomol Spectrosc *61*, 1529–1535.

Krafft, C., Knetschke, T., Funk, R.H.W., and Salzer, R. (2006). Studies on stress-induced changes at the subcellular level by Raman microspectroscopic mapping. Anal. Chem. *78*, 4424–4429.

Kroemer, G., Mariño, G., and Levine, B. (2010). Autophagy and the integrated stress response. Mol. Cell *40*, 280–293.

Kuntz, E., and Kuntz, H.-D. (2008). Hepatology: Textbook and Atlas (New York: Springer).

Landsberg, G., and Mandelstam, L. (1928). Eine neue Erscheinung bei der Lichtzerstreuung in Krystallen. Naturwissenschaften *16*, 557–558.

Lang, H., Sotiropoulos, G.C., Dömland, M., Frühauf, N.R., Paul, A., Hüsing, J., Malagó, M., and Broelsch, C.E. (2005). Liver resection for hepatocellular carcinoma in non-cirrhotic liver without underlying viral hepatitis. Br J Surg *92*, 198–202.

Laouirem, S., Le Faouder, J., Alexandrov, T., Mestivier, D., Léger, T., Baudin, X., Mebarki, M., Paradis, V., Camadro, J.-M., and Bedossa, P. (2014). Progression from cirrhosis to cancer is associated with early ubiquitin post-translational modifications: identification of new biomarkers of cirrhosis at risk of malignancy. J. Pathol. *234*, 452–463.

Le Faouder, J., Laouirem, S., Chapelle, M., Albuquerque, M., Belghiti, J., Degos, F., Paradis, V., Camadro, J.-M., and Bedossa, P. (2011). Imaging mass spectrometry provides fingerprints for distinguishing hepatocellular carcinoma from cirrhosis. J. Proteome Res. *10*, 3755–3765.

Lerou, P.H., and Daley, G.Q. (2005). Therapeutic potential of embryonic stem cells. Blood Rev. *19*, 321–331.

Li, X., Lin, J., Ding, J., Wang, S., Liu, Q., and Qing, S. (2004). Raman spectroscopy and fluorescence for the detection of liver cancer and abnormal liver tissue. Conf Proc IEEE Eng Med Biol Soc *1*, 212–215.

Llovet, J.M., Schwartz, M., and Mazzaferro, V. (2005). Resection and liver transplantation for hepatocellular carcinoma. Semin. Liver Dis. *25*, 181–200.

Lloyd, G.R., Wood, J., Kendall, C., Cook, T., Shepherd, N., and Stone, N. (2012). Histological imaging of a human colon polyp sample using Raman spectroscopy and self organising maps. Vibrational Spectroscopy *60*, 43–49.

Longuespée, R., Boyon, C., Castellier, C., Jacquet, A., Desmons, A., Kerdraon, O., Vinatier, D., Fournier, I., Day, R., and Salzet, M. (2012). The C-terminal fragment of the immunoproteasome PA28S (Reg alpha) as an early diagnosis and tumor-relapse biomarker: evidence from mass spectrometry profiling. Histochem. Cell Biol. *138*, 141–154.

Manghisi, G., Elba, S., Mossa, A., Giorgio, A., Aloisio, V., Perrotta, A., Tardio, B., and Del Naja, C. (1998). A new prognostic system for hepatocellular carcinoma: a retrospective study of 435 patients: the Cancer of the Liver Italian Program (CLIP) investigators. Hepatology *28*, 751–755.

Marvin, L.F., Roberts, M.A., and Fay, L.B. (2003). Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry in clinical chemistry. Clin. Chim. Acta *337*, 11–21.

Matthäus, C., Boydston-White, S., Miljković, M., Romeo, M., and Diem, M. (2006). Raman and infrared microspectral imaging of mitotic cells. Appl Spectrosc *60*, 1–8.

Mavarani, L., Petersen, D., El-Mashtoly, S.F., Mosig, A., Tannapfel, A., Kötting, C., and Gerwert, K. (2013). Spectral histopathology of colon cancer tissue sections by Raman imaging with 532 nm excitation provides label free annotation of lymphocytes, erythrocytes and proliferating nuclei of cancer cells. Analyst *138*, 4035–4039.

Mazzaferro, V., Regalia, E., Doci, R., Andreola, S., Pulvirenti, A., Bozzetti, F., Montalto, F., Ammatuna, M., Morabito, A., and Gennari, L. (1996). Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. N. Engl. J. Med. *334*, 693–699.

McDonnell, L.A., Heeren, R.M.A., Andrén, P.E., Stoeckli, M., and Corthals, G.L. (2012). Going forward: Increasing the accessibility of imaging mass spectrometry. J Proteomics *75*, 5113–5121.

McNeish, J. (2004). Embryonic stem cells in drug discovery. Nat Rev Drug Discov *3*, 70–80.

Meding, S., and Walch, A. (2013). MALDI imaging mass spectrometry for direct tissue analysis. Methods Mol. Biol. *931*, 537–546.

Meding, S., Balluff, B., Elsner, M., Schöne, C., Rauser, S., Nitsche, U., Maak, M., Schäfer, A., Hauck, S.M., Ueffing, M., et al. (2012). Tissue-based proteomics reveals FXYD3, S100A11 and GSTM3 as novel markers for regional lymph node metastasis in colon cancer. J. Pathol. *228*, 459–470.

Movasaghi, Z., Rehman, S., and Rehman, I.U. (2007). Raman Spectroscopy of Biological Tissues. Applied Spectroscopy Reviews *42*, 493–541.

Neugebauer, U., Clement, J.H., Bocklitz, T., Krafft, C., and Popp, J. (2010a). Identification and differentiation of single cells from peripheral blood by Raman spectroscopic imaging. Journal of Biophotonics *3*, 579–587.

Neugebauer, U., Clement, J.H., Bocklitz, T., Krafft, C., and Popp, J. (2010b). Identification and differentiation of single cells from peripheral blood by Raman spectroscopic imaging. J Biophotonics *3*, 579–587.

Neugebauer, U., Trenkmann, S., Bocklitz, T., Schmerler, D., Kiehntopf, M., and Popp, J. (2014). Fast differentiation of SIRS and sepsis from blood plasma of ICU patients using Raman spectroscopy. J Biophotonics *7*, 232–240.

Norris, J.L., and Caprioli, R.M. (2013). Analysis of Tissue Specimens by Matrix-Assisted Laser Desorption/Ionization Imaging Mass Spectrometry in Biological and Clinical Research. Chem Rev *113*, 2309–2342.

Oezdemir, R.F., Gaisa, N.T., Lindemann-Docter, K., Gostek, S., Weiskirchen, R., Ahrens, M., Schwamborn, K., Stephan, C., Pfister, D., Heidenreich, A., et al. (2012). Proteomic tissue profiling for the improvement of grading of noninvasive papillary urothelial neoplasia. Clin. Biochem. *45*, 7–11.

Oshima, Y., Shinzawa, H., Takenaka, T., Furihata, C., and Sato, H. (2010). Discrimination analysis of human lung cancer cells associated with histological type and malignancy using Raman spectroscopy. J Biomed Opt *15*, 017009.

Parkin, D.M., Bray, F., Ferlay, J., and Pisani, P. (2005). Global Cancer Statistics, 2002. CA: A Cancer Journal for Clinicians *55*, 74–108.

Patel, I.I., Trevisan, J., Evans, G., Llabjani, V., Martin-Hirsch, P.L., Stringfellow, H.F., and Martin, F.L. (2011). High contrast images of uterine tissue derived using Raman microspectroscopy with the empty modelling approach of multivariate curve resolution-alternating least squares. Analyst *136*, 4950–4959.

Poté, N., Alexandrov, T., Le Faouder, J., Laouirem, S., Léger, T., Mebarki, M., Belghiti, J., Camadro, J.-M., Bedossa, P., and Paradis, V. (2013). Imaging mass spectrometry reveals modified forms of histone H4 as new biomarkers of microvascular invasion in hepatocellular carcinomas. Hepatology *58*, 983–994.

Puskar, L., Tuckermann, R., Frosch, T., Popp, J., Ly, V., McNaughton, D., and Wood, B.R. (2007). Raman acoustic levitation spectroscopy of red blood cells and Plasmodium falciparum trophozoites. Lab Chip *7*, 1125–1131.

Raman, C.V., and Krishnan, K.S. (1928). A New Type of Secondary Radiation. Nature *121*, 501–502.

Rauser, S., Marquardt, C., Balluff, B., Deininger, S.-O., Albers, C., Belau, E., Hartmer, R., Suckau, D., Specht, K., Ebert, M.P., et al. (2010). Classification of HER2 receptor status in breast cancer tissues by MALDI imaging mass spectrometry. J. Proteome Res *9*, 1854–1863.

Rodrigo, M.A.M., Zitka, O., Krizkova, S., Moulick, A., Adam, V., and Kizek, R. (2014). MALDI-TOF MS as evolving cancer diagnostic tool: a review. J Pharm Biomed Anal *95*, 245–255.

Roskams, T., and Kojiro, M. (2010). Pathology of early hepatocellular carcinoma: conventional and molecular diagnosis. Semin. Liver Dis. *30*, 17–25.

Santiago-Cordoba, M.A., Romano, P.R., MacKay, A., and Demirel, M.C. (2011). Raman based hepatocellular carcinoma biomarker detection. Conf Proc IEEE Eng Med Biol Soc *2011*, 3672–3675.

Schmitt, M., and Popp, J. (2006). Raman spectroscopy at the beginning of the twenty-first century. J. Raman Spectrosc. *37*, 20–28.

Schröder, U.-C., Ramoji, A., Glaser, U., Sachse, S., Leiterer, C., Csaki, A., Hübner, U., Fritzsche, W., Pfister, W., Bauer, M., et al. (2013). Combined dielectrophoresis-Raman setup for the classification of pathogens recovered from the urinary tract. Anal. Chem. *85*, 10717–10724.

Schulze, H.G., Konorov, S.O., Piret, J.M., Blades, M.W., and Turner, R.F.B. (2013). Label-free imaging of mammalian cell nucleoli by Raman microspectroscopy. Analyst *138*, 3416–3423.

Short, K.W., Carpenter, S., Freyer, J.P., and Mourant, J.R. (2005). Raman spectroscopy detects biochemical changes due to proliferation in mammalian cell cultures. Biophys. J. *88*, 4274–4288.

Short, M.A., Lam, S., McWilliams, A., Zhao, J., Lui, H., and Zeng, H. (2008). Development and preliminary results of an endoscopic Raman probe for potential in vivo diagnosis of lung cancers. Opt Lett *33*, 711–713.

Smekal, A. (1923). Zur Quantentheorie der Dispersion. Naturwissenschaften *11*, 873–875.

Sugano, S., Miyoshi, K., Suzuki, T., Kawafune, T., and Kubota, M. (1994). Intrahepatic arteriovenous shunting due to hepatocellular carcinoma and cirrhosis, and its change by transcatheter arterial embolization. Am. J. Gastroenterol. *89*, 184–188.

Szaszák, M., Chang, J.C., Leng, W., Rupp, J., Ojcius, D.M., and Kelley, A.M. (2013). Characterizing the intracellular distribution of metabolites in intact Chlamydia-infected cells by Raman and two-photon microscopy. Microbes Infect. *15*, 461–469.

Taketani, A., Hariyani, R., Ishigaki, M., Andriana, B.B., and Sato, H. (2013). Raman endoscopy for the in situ investigation of advancing colorectal tumors in live model mice. Analyst *138*, 4183–4190.

Taleb, I., Thiéfin, G., Gobinet, C., Untereiner, V., Bernard-Chabert, B., Heurgué, A., Truntzer, C., Hillon, P., Manfait, M., Ducoroy, P., et al. (2013). Diagnosis of hepatocellular carcinoma in cirrhotic patients: a proof-of-concept study using serum micro-Raman spectroscopy. Analyst *138*, 4006–4014.

Tan, Y., Konorov, S.O., Schulze, H.G., Piret, J.M., Blades, M.W., and Turner, R.F.B. (2012). Comparative study using Raman microspectroscopy reveals spectral signatures of human induced pluripotent cells more closely resemble those from human embryonic stem cells than those from differentiated cells. Analyst *137*, 4509–4515.

User:Pavlina2.0, M. based on work of (2009). English: Molecular energy levels and Raman effect.

Végvári, Á., and Döme, B. (2011). State-of-the-art MS technology applications in lung disease. Bioanalysis *3*, 2665–2677.

Verrier, S., Notingher, I., Polak, J.M., and Hench, L.L. (2004). In situ monitoring of cell death using Raman microspectroscopy. Biopolymers *74*, 157–162.

Wang, Y.-J., Yao, H.-L., Wang, G.-W., Wang, Y., and Feng, M.-F. (2009). [Laser tweezers Raman spectroscopy analysis of liver cancer tissue]. Guang Pu Xue Yu Guang Pu Fen Xi *29*, 1881–1883.

Wiley, W.C., and McLaren, I.H. (1955). Time-of-Flight Mass Spectrometer with Improved Resolution. Review of Scientific Instruments *26*, 1150–1157.

Wilson, L., and Jordan, M.A. (1995). Microtubule dynamics: taking aim at a moving target. Chem. Biol. *2*, 569–573.

Wu, Y.-M., Chen, H.-C., Chang, W.-T., Jhan, J.-W., Lin, H.-L., and Liau, I. (2009). Quantitative Assessment of Hepatic Fat of Intact Liver Tissues with Coherent Anti-Stokes Raman Scattering Microscopy. Anal. Chem. *81*, 1496–1504.

Xiong, Y., Li, Y.-T., Guo, Y., Rao, F.-F., Zhang, J.-M., Si, M.-Z., Liu, R.-M., and Tang, W.-Y. (2012). [Analysis of surface enhanced Raman scattering spectra of oxyhemoglobin for liver cancer with combined multivariate statistics]. Guang Pu Xue Yu Guang Pu Fen Xi *32*, 2427–2432.

Yahagi, N., Shimano, H., Hasegawa, K., Ohashi, K., Matsuzaka, T., Najima, Y., Sekiya, M., Tomita, S., Okazaki, H., Tamura, Y., et al. (2005). Co-ordinate activation of lipogenic enzymes in hepatocellular carcinoma. Eur. J. Cancer *41*, 1316–1322.

Yalcin, E.B., and de la Monte, S.M. (2015). Review of Matrix-Assisted Laser Desorption Ionization-Imaging Mass Spectrometry for Lipid Biochemical Histopathology. J Histochem Cytochem.

Yu, C., Gestl, E., Eckert, K., Allara, D., and Irudayaraj, J. (2006). Characterization of human breast epithelial cells by confocal Raman microspectroscopy. Cancer Detect. Prev. *30*, 515–522.

Zhu, K., Dai, Z., and Zhou, J. (2013). Biomarkers for hepatocellular carcinoma: progression in early diagnosis, prognosis, and personalized therapy. Biomarker Research *1*, 10.

Zoladek, A., Pascut, F.C., Patel, P., and Notingher, I. (2011). Non-invasive time-course imaging of apoptotic cells by confocal Raman micro-spectroscopy. J. Raman Spectrosc. *42*, 251–258.

Curriculum vitae

Personal Data

Name	Tolstik Tatiana
Date of birth	July 6th, 1987
Place of birth	Minsk, Belarus
E-mail	tolstiktatiana@gmail.com

Education

Since August 2015	**Postdoc** at Center for Sepsis Control and Care at the Jena University Hospital
2010-2015	**PhD-student** at the Department of Internal Medicine IV, Division of Gastroenterology, Hepatology and Infectious Diseases in **Jena University Hospital** and **Leibniz Institute of Photonic Technology, Jena**
2005 – 2010	**Diploma** in Biology with specialization in Genetics and Teaching degree of Biology and Chemistry at the Belarusian State University, Minsk, Graduated from Faculty of Biology, Department of Genetics [GPA[§] 10]
2003 – 2005	Lyceum of Belarusian State University; Graduation from biologic-chemical class with high average. [GPA[§] 8]
1994 – 2003	School № 191, Minsk; Graduation from mathematical-informatical class with high average [GPA[§] 9]

Practical Experience

2005 – 2009	Research assistant at the Faculty of Biology, Department of Genetics at the Belarusian State University of Minsk, Belarus
June- August 2008	Research internship at the Belarusian State University Minsk, Belarus [GPA[§] 10]
February 2009	Teaching practicum in biology at the School № 52, Minsk, Belarus
June- August 2009	Research internship during Bachelor study at the Friedrich-Schiller-University Jena, Germany, Institute of Human Genetics and Anthropology, Laboratory of Molecular Genetics [GPA[§] 9]
March-June 2010	Research internship at the Friedrich-Schiller-University Jena, Germany, Laboratory of Organic and Macromolecular Chemistry [GPA[§] 10]

[§] GPA = Grade Point Average: 10-9 = excellent, 8 = very good, 7-6 = good, 5-4 = satisfactory

Publication list

Scientific publications

1. T. Tolstik, C. Marquardt, C. Beleites, C. Matthäus, C. Bielecki, M. Bürger, C. Krafft, O. Dirsch, U. Settmacher, J. Popp, A. Stallmach “Classification and prediction of HCC tissues by Raman imaging with identification of fatty acids as potential lipid biomarkers”, *J Cancer Res Clin Oncol*, **2015 Mar**; 141(3): 407-18
2. C. Marquardt, T. Tolstik, C. Bielecki, R. Kaufmann, A. Crecelius, U.S. Schubert, U. Settmacher, A. Stallmach, O. Dirsch “MALDI imaging-based classification of hepatocellular carcinoma and non-malignant lesions in fibrotic liver tissue”, *Z Gastroenterol* **2015 Jan**; 53(1): 33–39
3. T. Tolstik, C. Marquardt, C. Matthäus, N. Bergner, C. Bielecki, C. Krafft, A. Stallmach, J. Popp “Discrimination and Classification of Liver Cancer Cells and Proliferation States by Raman Spectroscopic Imaging”, *Analyst*, **2014 Nov**; 139(22): 6036-43
4. K. Babiuch, D. Pretzel, T. Tolstik, A. Vollrath, S. Stanca, F. Foertsch, C. R. Becer, M. Gottschaldt, C. Biskup, U. S. Schubert “Uptake of Well-Defined, Highly Glycosylated, Pentafluorostyrene-Based Polymers and Nanoparticles by Human Hepatocellular Carcinoma Cells”, *Macromol Biosci* **2012 Sep**; 12(9): 1190-9

Oral presentations

1. **April 2015**, Diagnostic by Raman spectroscopic imaging of the patients with Ulcerative colitis and anti-TNF antibody therapy for prediction of a stable and long-lasting clinical remission, *Multi-Field Clinic of XXI Century*, St. Petersburg, Russia
2. **October 2014**, Raman based diagnostics of IBD, *Forum Intestinale Immunologie*, Berlin, Germany
3. **October 2014,** New molecular classification models based on mass spectrometry and Raman spectroscopic imaging for improved prognostic evaluation of liver cancer, *Clinical Group Seminar Heptology and Infectiology,* Jena, Germany
4. **September 2014,** Laureate lecture, *Forschungspreis der DCCV 2014*, Leipzig, Germany

5. **September 2014**, Raman spectroscopic imaging for investigations of de novo lipogenesis in hepatocellular carcinoma, G*erman Society for Gastroenterology, Digestive and Metabolic Diseases*, Leipzig, Germany
6. **Mai 2014,** Raman spectroscopic imaging for investigations of lipogenesis in hepatocellular carcinoma, *IPHT Group Seminar*, Jena, Germany
7. **April 2014**, Raman spectroscopy – new diagnostic method for hepatocellular carcinoma, *Multi-Field Clinic of XXI Century. High-Tech Medical Assistance*, St. Petersburg, Russia
8. **March 2014,** Clinical applications of Raman spectroscopy, *P4L Interdisciplinary School on Clinical Biophotonics*, Jena, Germany
9. **February 2014,** New molecular classification models based on mass spectrometry and Raman spectroscopic imaging for improved prognostic evaluation of liver cancer, *Dornburg Seminar*, Dornburg, Germany
10. **April 2013,** Diagnostics of Primary Liver Cancer, *IPHT Group Seminar*, Jena, Germany
11. **December 2012**, Raman spectroscopic imaging of liver tissue, *IPHT Group Seminar*, Jena, Germany
12. **November 2012,** Raman spectroscopy for assessment of liver cells and tissue, *Dornburg Seminar*, Dornburg, Germany
13. **Mai 2012,** Liver cancer and Raman spectroscopic imaging, *IPHT Group Seminar*, Jena, Germany

Posters in medical conferences

1. T. Tolstik, C. Marquardt, C. Matthäus, C. Beleites, C. Bielecki, C. Krafft, O. Dirsch, J. Popp, A. Stallmach “Molecular Investigations by Raman Imaging for a Reliable Diagnosis of Hepatocellular Carcinoma”, *UEG Week 2013*, Berlin, Germany, **2013**
2. T. Tolstik, C. Marquardt, C. Matthäus, C. Krafft, C. Bielecki, O. Dirsch, J. Popp, A. Stallmach “Classifying malignant transformed hepatocytes by Raman imaging”, *UEG Week 2012*, Amsterdam, The Netherlands, **2012**

3. C. Marquardt, T. Tolstik, C. Bielecki, R. Kaufmann, A. Crecelius, U.S. Schubert, U. Settmacher, A. Stallmach, O. Dirsch “Molekulare Differenzierung von malignen und nicht-malignen Läsionen in Kryogewebeschnitten des Hepatozellulären Karzinoms durch MALDI IMS“, *55. Jahrestagung der Gesellschaft für Innere Medizin Thüringens e.V.*, Jena, Germany, **2011**

Posters in spectroscopy conferences

1. T. Tolstik, C. Marquardt, C. Matthäus, C. Beleites, C. Krafft, C. Bielecki, O. Dirsch, U. Settmacher, A. Stallmach, J. Popp “Raman imaging spectroscopy for investigations of de novo lipogenesis in hepatocellular carcinoma”, *CTCT2015: Current Trends in Cancer Theranostics*, Jena, Germany, **2015**.
2. T. Tolstik, C. Marquardt, C. Matthäus, C. Beleites, C. Bielecki, C. Krafft, O. Dirsch, U. Settmacher, J. Popp, A. Stallmach “Investigations of de novo lipogenesis in hepatocellular carcinoma by Raman spectroscopic imaging”, *11th Symposium Confocal Raman Imaging*, Ulm, Germany, **2014**
3. T. Tolstik, C. Marquardt, C. Matthäus, C. Beleites, C. Bielecki, C. Krafft, O. Dirsch, U. Settmacher, J. Popp, A. Stallmach “New molecular classification models based on Raman spectroscopic imaging for improved prognostic evaluation of liver cancer”, *24th International Conference on Raman Spectroscopy*, Jena, Germany, **2014**
4. T. Tolstik, C. Marquardt, C. Matthäus, C. Beleites, C. Bielecki, C. Krafft, O. Dirsch, J. Popp, A. Stallmach “Random Forest Classification of Raman Spectra for Reliable Diagnosis of Hepatocellular Carcinoma”, *FT-IR Workshop*, Berlin, Germany, **2013**

Awards

1. **September 2014**, an abstract preis at the conference Die Deutsche Gesellschaft für Gastroenterologie, Verdauungs- und Stoffwechselkrankheiten, Leipzig, Germany
2. **August 2014,** Hermann-Strauß-Forschungspreis der DCCV 2014 donated by the company AbbVie Germany GmbH & Co. KG, Germany
3. **October 2013**, “ProChance 2013” for the Promotion of Equal Opportunities for Women and Men in Science, Programme Line A 2, Friedrich-Schiller University Jena, Germany

4. **October 2012**, “ProChance 2012” for the Promotion of Equal Opportunities for Women and Men in Science, Programme Line A 2, Friedrich-Schiller University Jena, Germany

Acknowledgement

I am using this opportunity to express my gratitude to everyone who has contributed to the work described in the presented thesis.

First of all, I would like to thank my promoter, Prof. Andreas Stallmach for giving me the opportunity to work in his group at the department of Gastroenterology, Hepatology and Infectious Diseases and for his important and continuous support. Dear Prof. Stallmach, thank you for encouraging me and for your personal guidance. I am grateful for all possibilities and freedom that you gave me while doing my research. You provided me the best environment for fruitful collaborations within and outside the group.

I would like to thank my second promoter, Prof. Jürgen Popp for the opportunity to join his group at Leibniz Institute of Photonic Technology, Jena and to become a part of it. Especially, I am thankful to the "Optical Cell Diagnostics" group for warm welcome and all support they provided me. Dear Prof. Popp, I have learned a lot in your group.

I would like to express my gratitude to my supervisor M.Sc. Claudio Marquardt for helping me to get accepted as a PhD student at the Medical Faculty. Thank you Claudio, for your aspiring guidance, introduction to the field of Maldi Imaging, constructive criticism and help in manuscripts preparation. I am grateful to Dr. med. Tony Bruns and all colleges from division of Gastroenterology, Hepatology and Infectious Diseases, University hospital Jena for sharing their truthful and illuminating views on our projects and possibility to learn a lot about medicine. Thank you, dear Marita Vetterlein, for your help, organization of the Gastro Lab and excellent technical assistance. I want to express my warm thanks to Dr. med. Christiane Bielecki for her friendship, succesfull discussions and first Raman experiments that we did together.

I would like to thank Dr. Christian Mathaeus for his support, guidance and friendly advice during the Raman spectroscopy project work. Thank you Christian, for providing me help with all ideas and projects and teaching me a lot about Raman spectroscopy. Our specialists in chemometrics, Dr. Claudia Beleites, Dr. Thomas Bocklitz and Dr. Norbert Bergner, thank you so much for all time you spend for our projects and all things you taught me. I would like to thank especially our offices in IPHT and group fun time together, people with whom I shared every day of this PhD time.

At the end, I would like to thank my family who raised me with the faith in science. Thank you for your support and encouraging me in realization of my dreams and all help during my PhD. Especially I am grateful to my sister Elena who was the reason for me to come to Jena and helped me to start my life in Germany. Thank you for always been there for me. Biggest thanks goes to my mother and father, you have dedicated your life for our education and perspective future. Thank you dear Teresia Hallstrom for your energy, friendship and all advices you gave me in research and life. And, of course, I am thankful to all my friends and close people who supported me even at a distance.

Thank you all!

Selbständigkeitserklärung

Hiermit erkläre ich, dass mir die Promotionsordnung der Medizinischen Fakultät der Friedrich-Schiller-Universität bekannt ist,

ich die Dissertation selbst angefertigt habe und alle von mir benutzten Hilfsmittel, persönlichen Mitteilungen und Quellen in meiner Arbeit angegeben sind,

dass Personen der Auswahl und Auswertung des Materials sowie der Herstellung des Manuskripts unterstützt haben, vollständig erklärt,

die Hilfe eines Promotionsberaters nicht in Anspruch genommen wurde und dass Dritte weder unmittelbar noch mittelbar geldwerte Leistungen von mir für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen,

dass ich die Dissertation noch nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht habe und

dass ich die gleiche, eine in wesentlichen Teilen ähnliche oder eine andere Abhandlung nicht bei einer anderen Hochschule als Dissertation eingereicht habe.

Die Ergebnisse der Publikation "Klassifikation des hepatozellulären Karzinoms und nicht-maligner Läsionen im fibrotischen Lebergewebe mittels bildgebender Massenspektrometrie" in Master Thesis von Claudio Marquardt eingereicht an der Medizinischen Fakultät der Friedrich-Schiller-Universität verwendet worden. Mein Anteil an diesem Paper waren die H&E-Färbungen der Gewebe (zu ca. 40 %) mit dem anschließenden Scannen aller gefärbten Objektträger. Danach erfolgte die Bildbearbeitung (co-registration) für alle Proben und die Datenanalyse (zu ca. 30 %). Aufgrund der beschriebenen Beteiligung an der Gesamtarbeit, begründet dies die zweite Position als Autor.

Jena, 5. November 2015

Declaration of authorship

I declare in lieu of oath that I have read doctoral regulations of the Medical Faculty of the Friedrich Schiller University,

the work presented here is, to the best of my knowledge and belief, original and the result of my own investigations, except as acknowledged,

that persons, who have supported in the selection and assessment of materials and in the preparation of the manuscript, are declared completely,

the assistance of doctoral supervisor has not been utilizes and no third parties have either direct or indirect received monetary benefits from the candidate for work related to the submitted Doctoral thesis,

that this dissertation has not yet been submitted, either in part or whole, as an examination paper for a state or other academic examinations.

The results of publication "MALDI Imaging-based Classification of Hepatocellular Carcinoma and Non-Malignant Lesions in Cirrhotic Liver Tissue" were used in Master Thesis of Claudio Marquardt submitted at the Medical Faculty of the Friedrich Schiller University. My personal input in this article was H&E staining of tissue sections (for ca. 40%) with subsequent scanning of all slides. Followed by performing image co-registeration for all tissue specimens and data analysis (for ca. 30 %). Described input to this research work resulted in second authorship position.

Jena, 5 November 2015